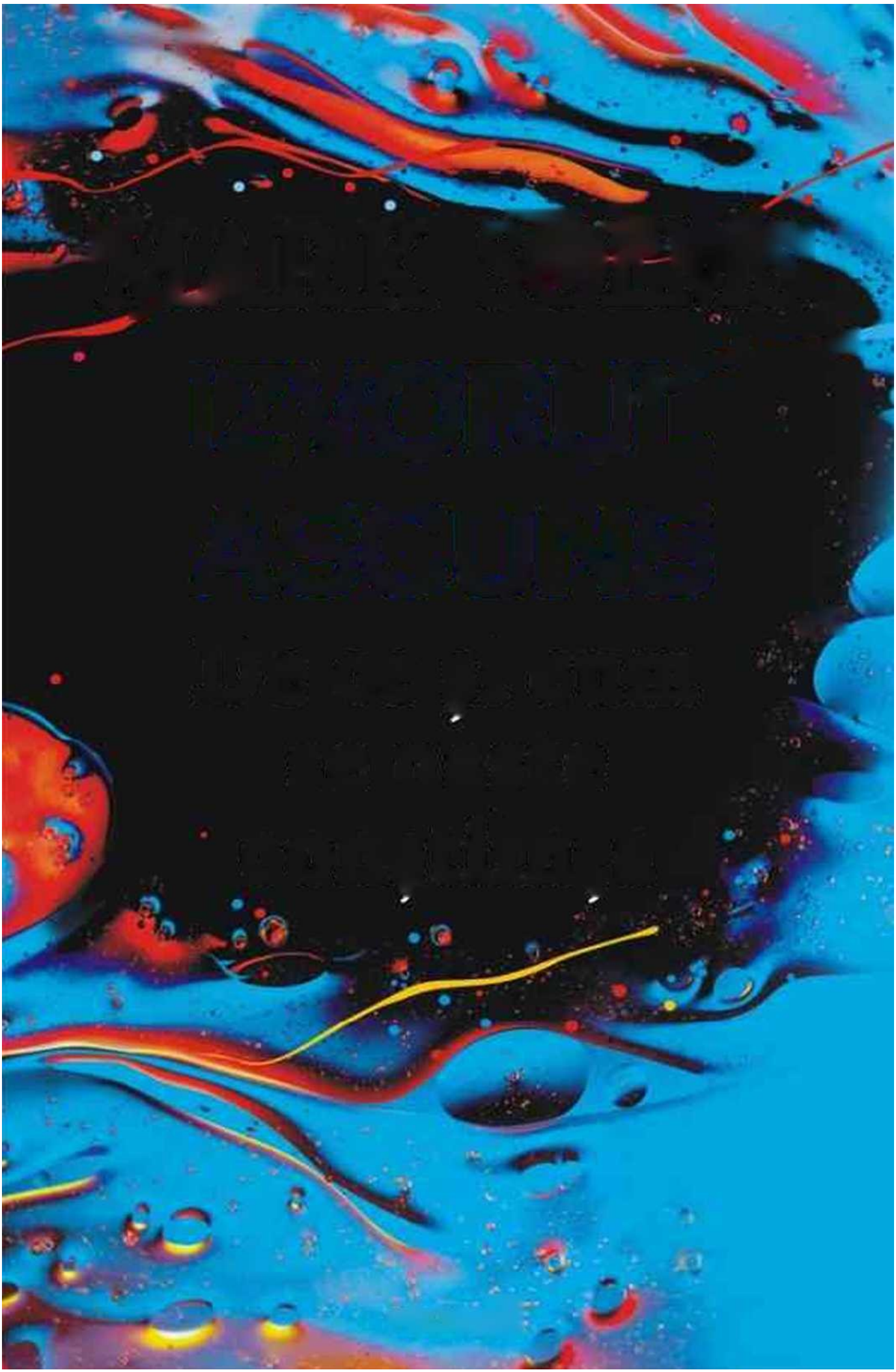


MARK SOLMS

IZVORUL ASCUNS

De ce şi cum se naşte conştiinţa

Traducere de Dan Bălănescu

EDITURA POLIROM
2022

MARK SOLMS, psihanalist şi neuropsiholog, a descoperit mecanismele cerebrale ale viselor. Este cercetător în cadrul departamentului de neuropsihologie de la Neuroscience Institute, University

of Cape Town. În anul 2000 a înfiinţat International Neuropsychoanalysis Society. Este conferenţiar onorific în neurochirurgie al Royal London Hospital School şi membru de onoare al American College of Psychiatrists. A mai publicat *The Brain and the Inner World* (împreună cu Oliver Turnbull) (2002).

În amintirea lui
Jaak Panksepp
(1943-2017)

A rezolvat vechea ghicitoare şi a fost un om foarte înţelept.

Lista figurilor

1. Principalele structuri cerebrale implicate în somnul REM.
2. Principalele structuri cerebrale implicate în vis
3. Scanări PET ale creierului la debutul somnului şi în timpul somnului REM
4. RMN al unui copil născut fără cortex cerebral (Copyright © ACR Learning File, Neuroradiology, Edition 2, 2004, © American College of Radiology)
5. Reacţie emoţională a unui copil născut fără cortex cerebral (figură din „Consciousness without a cerebral cortex: A challenge for neuroscience and medicine” de Bjorn Merker, *Behavioral and Brain Sciences,* 1 mai 2007, Cambridge University Press © 2007 Cambridge University Press)
6. Coliculii superiori, cortexul vizual şi PAG
7. RMN-uri ale unui bărbat prezentând distrugere completă a lobilor prefrontali
8. Prima diagramă a sistemelor memoriei realizată de Freud
9. Imagistică RMN a creierului în patru stări emoţionale
10. Tipare tipice ale undelor activităţii corticale înregistrate cu ajutorul EEG

11. Reprezentare grafică a măsuritorilor potenţialelor de acţiune *(spike trains)* a 20 de neuroni (figură din „Simulating Neural Spike Trains” de Praneeth Namburi, Praneethnamburi.com, 5 februarie 2015 © Praneeth Namburi 2015)
12. Homeostaza sentimentului
13. Sistem cu autoorganizare şi zona sa Markov (figură din „Life
    aş We Know It” de Karl Friston, *Journal of the Royal Society Interface,* 10,6 septembrie 2013, https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0475, CC BY)
14. Efecte entropice ale unor leziuni la nivelul zonei Markov
15. Ierarhie predictivă simplificată
16. Reprezentare grafică a unei predicţii posterioare (figură din Hidesyoshi Yanagisawa, Oto Kawamata şi Kazutaka Ueda, „Modeling Emotions Associated with Novelty at Variable Uncertainty Levels: A Bayesian Approach”, *Frontiers in Computational Neuroscience,* 24 ianuarie 2012, https://www.frontiersin.org/articles/10.3389/fncom.2019.00002/full)
17. Dinamica unui sistem care se autovalidează, cu optimizarea preciziei
18. Testul figurilor complexe Rey aplicat unui copil cu

vedere inversată

19. Schemă de învăţare prin eroare de predicţie a recompensei

Introducere

Când eram mic, mi-am pus o întrebare ciudată: cum ne putem imagina lumea care exista înainte de evoluţia conştiinţei? Desigur, a existat o astfel de lume, dar cum ţi-o *imaginezi* – lumea aşa cum era înainte ca reprezentarea lucrurilor să devină posibilă?

Ca să înţelegi ce vreau să spun, încearcă să-ţi imaginezi o lume în care *un răsărit de soare nu poate avea loc.* Pământul s-a rotit întotdeauna în jurul Soarelui, dar soarele răsare la orizont doar din punctul de vedere al unui observator. Este un eveniment care ţine inerent de perspectivă. Răsăritul soarelui va fi întotdeauna prins în capcana experienţei subiective.

Această asumare obligatorie a perspectivei este ceea ce face atât de dificilă înţelegerea conştiinţei. Dacă vrem să reuşim, trebuie să eludăm subiectivitatea – să privim conştiinţa din exterior, să vedem lucrurile aşa cum sunt cu adevărat, şi nu cum ni se par. Dar cum facem asta? Cum ne desprindem de propriul sine?

În tinereţe, îmi vizualizam cu naivitate conştiinţa ca pe o bulă care mă înconjoară: conţinutul ei era format din imaginile în mişcare, sunetele şi alte fenomene ale experienţei. Dincolo de această bulă, presupuneam că

era un întuneric infinit. Îmi imaginam acest întuneric ca pe o simfonie a mărimilor adimensionale, a forţelor interdependente, a energiilor şi a altora asemenea: adevărata realitate „din exterior” pe care conştiinţa mea o reprezintă în formele calitative inevitabile.

Imposibilitatea oricărei astfel de imaginări – a reprezentării realităţii fără reprezentări – ilustrează amploarea sarcinii pe care mi-am propus-o în această carte. Încă o dată, după atâţia ani, încerc să arunc o privire în spatele vălului conştiinţei, să-i surprind mecanismele.

Din acest motiv, cartea de faţă e inevitabil perspectivală. De fapt, e chiar mai perspectivală decât i-o impune paradoxul pe care tocmai l-am descris. Ca să te ajut să vezi lucrurile din punctul meu de vedere, m-am hotărât să-ţi împărtăşesc o parte din povestea mea. Progresele în ideile mele ştiinţifice despre conştiinţă au apărut adesea în urma unor evenimente din viaţa mea personală şi din activitatea mea clinică şi, deşi cred că concluziile mele sunt de sine stătătoare, e mult mai uşor să le înţelegi dacă ştii cum am ajuns la ele. Unele dintre descoperirile mele – de exemplu, mecanismele cerebrale ale visului – au fost în mare parte întâmplătoare. Unele dintre alegerile mele din punct de vedere profesional – de exemplu, să mă abat puţin de la cariera mea în neuroştiinţe şi să mă formez ca psihanalist – mi-au adus mai multe satisfacţii decât aş fi putut spera. În ambele cazuri, voi explica şi cum.

Dar, în măsura în care misiunea mea de a înţelege conştiinţa şi-a atins scopul, adevăratul meu noroc este

că am avut nişte colaboratori geniali. În special, am avut marele noroc să lucrez cu regretatul Jaak Panksepp, un specialist în neuroştiinţe care, mai mult decât oricare altul, a înţeles originea şi puterea *sentimentelor.* Cam tot ce cred eu astăzi despre creier a fost influenţat de ideile lui.

Mai recent, am avut posibilitatea să lucrez cu Karl Friston, care, printre numeroasele sale calităţi formidabile, are şi renumele de a fi cel mai influent specialist în neuroştiinţe în viaţă din lume. Friston a pus bazele cele mai solide ale teoriei pe care mă pregătesc să o elaborez. E celebru în principal pentru că a redus funcţiile cerebrale (de toate tipurile) la o necesitate fizică fundamentală de a minimiza ceva numit *energie liberă.* Explic acest concept în capitolul 7, dar, deocamdată, voi preciza doar că teoria la care am lucrat împreună cu Friston se leagă de acest proiect – în asemenea măsură încât poate fi numită teoria energiei libere a conştiinţei. Despre asta e vorba.

Explicaţia finală a capacităţii de a simţi este o enigmă atât de complicată, încât astăzi este numită cu respect „problema dificilă”. Uneori, după rezolvarea unei enigme, atât întrebarea, cât şi răspunsul încetează să mai fie interesante. Rămâne la latitudinea ta să stabileşti dacă ideile pe care le prezint aici aruncă o lumină nouă asupra acestei probleme dificile. În orice caz, sunt convins că te vor ajuta să te vezi pe *tine însuţi* într-o lumină nouă şi, în acest sens, ar trebui să rămână interesante până când sunt schimbate. Până la urmă, într-un sens profund, tu *eşti* conştiinţa ta. Aşadar, pare

raţional să te aştepţi ca o teorie a conştiinţei să explice fundamentele motivelor pentru care simţi aşa cum simţi. Ar trebui să explice de ce eşti aşa cum eşti. Poate că ar trebui să clarifice chiar şi ce poţi face în această privinţă.

Recunosc că ultima temă trece dincolo de ce mi-am propus legat de sfera acestei cărţi. Dar nu şi dincolo de sfera teoriei. Ceea ce spun eu despre conştiinţă reuneşte într-o singură poveste fizica elementară a vieţii, cele mai recente progrese în neuroştiinţa computaţională şi în cea afectivă, precum şi subtilităţile experienţei subiective analizate de obicei de psihanalişti. Cu alte cuvinte, lămuririle aduse de această teorie ar trebui să-ţi fie folositoare.

Această teorie e munca mea de-o viaţă. După multe decenii, încă mă întreb cum ar fi putut arăta lumea înainte să fi existat cineva care să o vadă. Acum, că deţin mai multă cunoaştere, îmi imaginez zorii vieţii într-un izvor geotermal. Organismele unicelulare ce au luat fiinţă acolo cu siguranţă nu au fost conştiente, dar şansele lor de supravieţuire au fost afectate de mediul ambiant. E uşor să ne imaginăm aceste organisme simple răspunzând „bunăvoinţei” biologice a energiei solare. De acolo, nu mai e decât un pas până la a ne imagina creaturi mai complexe care caută activ asemenea surse de energie şi care, în cele din urmă, dezvoltă o capacitate de a evalua şansele de succes prin acţiuni alternative.

În opinia mea, conştiinţa a luat naştere din experienţa unor asemenea organisme. Imaginează-ţi

arşiţa zilei şi frigul nopţii din perspectiva acestor prime fiinţe. Valorile fiziologice care le înregistrau experienţele diurne au fost precursoarele primului răsărit de soare.

Mulţi filosofi şi oameni de ştiinţă încă cred că simţirea nu serveşte niciunui scop fizic. Sarcina mea în această carte este să te conving că o interpretare alternativă este plauzibilă. Asta mă obligă să te conving că sunt parte a naturii, că nu sunt fundamental diferite de alte fenomene naturale şi că ele *fac* ceva în interiorul matricii cauzale a lucrurilor. Voi demonstra că conştiinţa este legată de faptul de a simţi, iar faptul de a simţi, la rândul său, are legătură cu cât de bine sau de prost te descurci în viaţă. Conştiinţa există ca să te ajute să te descurci
mai bine.

Se spune că problema dificilă a conştiinţei este cea mai mare enigmă nerezolvată a neuroştiinţelor contemporane, dacă nu chiar a ştiinţei în general. Soluţia propusă în această carte se îndepărtează radical de abordările convenţionale. Întrucât cortexul cerebral este sediul inteligenţei, aproape toate lumea crede că e şi sediul conştiinţei. Eu nu sunt de acord – conştiinţa este mult mai primitivă. Ea apare într-o zonă a creierului pe care o au şi oamenii, şi peştii. Acesta este „izvorul ascuns” din titlu.

Conştiinţa nu trebuie confundată cu inteligenţa. Poţi să simţi durere fără să ai vreo idee despre cauza ei. De asemenea, nu e neapărat ca impulsul de a mânca –

senzaţia de foame – să implice vreo comprehensiune intelectuală a exigenţelor vieţii. Conştiinţa în forma ei elementară, şi anume *sentimentele* brute, este o funcţie surprinzător de simplă.

Regăsim această abordare şi la alţi trei specialişti de marcă în neuroştiinţe: Jaak Panksepp, Antonio Damasio şi Bjorn Merker. Panksepp a fost deschizător de drumuri. El (ca şi Merker) a făcut cercetări pe animale, Damasio (ca şi mine) nu. Mulţi cititori se vor îngrozi de concluziile studiilor pe animale pe care le voi prezenta aici, tocmai deoarece demonstrează că alte animale simt la fel ca noi. Toate mamiferele simt durere, frică, panică, tristeţe şi altele asemenea. E o ironie că tocmai cercetările lui Panksepp au eliminat orice urmă de îndoială în această privinţă. Singura noastră consolare este că descoperirile lui au făcut imposibilă continuarea în aceeaşi manieră a unor asemenea studii.

Ideile lui Panksepp, Damasio şi Merker m-au atras deoarece aceştia cred, ca şi mine, că neurostiintelor actuale le lipseşte o focalizare clară pe natura concretă a *experienţei trăite.* Se poate spune că ceea ce ne uneşte este că am clădit, uneori fără să ne dăm seama, pe temeliile abandonate pe care le-a turnat Freud pentru o ştiintă a gândirii care pune simţămintele înaintea cognitiei. (Cognitia este în cea mai mare parte inconştientă.) Acesta e al doilea sens în care volumul de fată se desprinde radical de teoriile general acceptate – se întoarce la „Proiectul” lui Freud din 1895 şi încearcă să-i duc sarcina la bun sfârşit. Dar nu îi trec cu vederea numeroasele erori. De exemplu, ca toată lumea, Freud

credea că conştiinta este o functie corticală.

În al treilea rând, volumul de fată ajunge la concluzia că conştiinta se pretează la inginerii. Că poate fi *produsă* artificial. Această concluzie, cu profundele sale implicatii metafizice, izvorăşte din colaborarea mea cu Karl Friston. Spre deosebire de Panksepp, Damasio şi Merker, Friston este specialist în neuroştiintele computationale. Aşadar, el crede că conştiinta este în esenţă reductibilă la legile fizicii (convingere în mod surprinzător împărtăşită de Freud). Însă chiar şi Friston a pus în mare parte semnul egalitătii între functiile psihice şi cele corticale înainte să ne începem colaborarea. Această carte duce cadrul statistico-mecanic al lui Friston la un nivel mai profund, în cele mai primitive cotloane ale trunchiului cerebral...

Aceste trei perspective diferite fac problema dificilă mai putin dificilă. Iar cum anume va explica volumul de fată.

Mark Solms
Chailey, East
Sussex martie
2020

Materia viselor

M-am născut pe Coasta Scheletelor din fosta colonie germană Namibia, unde tatăl meu administra o mică firmă sud- africană numită Consolidated Diamond Mines. Compania- mamă, De Beers, crease practic un

stat în stat, numit *Sperrgebiet* („zonă interzisă”). Minele sale aluvionare se întindeau de la dunele de nisip ale Deşertului Namib până pe fundul Oceanului Atlantic, pe o distanţă de mai mulţi kilometri în larg.

Acesta a fost peisajul aparte care mi-a modelat imaginaţia. În copilărie, mă jucam cu fratele meu mai mare Lee de-a minele de diamante, folosind excavatoare de jucărie, recreând în grădina noastră impresionantele realizări inginereşti la care eram martori când ne ducea tata să vedem exploataţiile miniere de suprafaţă din deşert. (Eram, desigur, prea mici ca să cunoaştem aspectele mai puţin impresionante ale ocupaţiei lui tata.)

Într-o zi din 1965, când aveam patru ani, părinţii mei se plimbau cu iahtul la Cormorant Yacht Club, cum făceau adesea, iar eu mă jucam în club cu Lee, care avea şase ani. Soarele arzător risipise ceaţa de la primele ore ale dimineţii. Am ieşit

din interiorul răcoros al clădirii de trei etaje a clubului, îndreptându-mă spre malul apei. Mă plimbam cu picioarele în apă şi urmăream nişte peştişori sclipitori cum se îndepărtează de degetele mele, în timp ce Lee şi nişte prieteni de-ai lui s-au căţărat pe acoperiş.

Am trei flashbackuri. Primul, cu sunet ca pârâitul unui pepene verde care se crapă. Al doilea, imaginea lui Lee întins la pământ şi smiorcăindu-se că îl doare un picior. Ultimul, mătuşa şi unchiul spunându-mi că vor avea ei grijă de mine şi de sora mea cât părinţii noştri merg la spital cu Lee. Partea cu piciorul dureros e probabil o închipuire: în dosarul medical scrie că fratele meu şi-a pierdut cunoştinţa la impactul cu trotuarul de beton.

Lee avea nevoie de un tip de îngrijiri de specialitate pe care spitalul nostru local nu le putea oferi. A fost dus cu elicopterul la spitalul Groote Schuur din Cape Town, la 800 de kilometri distanţă. Pe vremea aceea, secţia de neurochirurgie funcţiona într-o clădire impresionantă construită în stil Cape Dutch, exact aceeaşi clădire în care lucrez eu acum ca neuropsiholog. Lee îşi fracturase craniul şi suferise o hemoragie intracraniană. Când se extind astfel de hematoame, ele constituie o urgenţă ce ameninţă viaţa şi necesită intervenţie chirurgicală. Fratele meu a avut noroc: şi-a revenit în următoarele câteva zile şi, în cele din urmă, a fost externat şi trimis acasă.

În afară de faptul că după accident a trebuit să poarte o cască de protecţie pentru craniul fracturat, Lee nu arăta diferit. Totuşi, persoanalitatea lui se modificase profund. Germana are un cuvânt pentru senzaţia pe

care mi-a provocat-o acest lucru, *Unheimlichkeit,* pentru care nu există un echivalent exact în engleză. Literal, înseamnă „ciudăţenie”, dar s-ar traduce mai bine prin „ceva care îţi dă fiori reci” sau „stranietate”.

Cel mai evident, s-a schimbat în sensul că şi-a pierdut achiziţiile ce marchează trecerea într-o nouă etapă de dezvoltare. O vreme, şi-a pierdut parţial chiar şi controlul sfincterelor. Ceea ce m-a tulburat mai mult a fost faptul că părea să *gândească* altfel decât înainte. Parcă Lee era şi în acelaşi timp nu era prezent. Parcă uitase multe dintre jocurile pe care le jucam. Acum, jocul nostru de-a săpatul după diamante a devenit pur şi simplu săpat gropi. Aspectele imaginative şi simbolice ale jocului nu îl mai atrăgeau. Nu mai era Lee.

Nu a trecut clasa – era primul său an de şcoală. Din acele prime zile după accident îmi amintesc mai des încercarea mea de a mă împăca cu dihotomia că fratele meu care se întorsese acasă arăta la fel, dar nu era la fel. Mă întrebam unde dispăruse versiunea sa anterioară.

În anii ce au urmat am căzut în depresie. Îmi amintesc că nu reuşeam să mă mobilizez ca să mă încalţ dimineaţa, să merg la şcoală. Asta se întâmpla la vreo trei ani după accident. Nu îmi găseam energia să fac aceste lucruri pentru că nu le mai vedeam rostul. Dacă însăşi fiinţa noastră depinde de funcţionarea creierului, oare ce mi se va întâmpla când va muri creierul *meu,* ce se va întâmpla cu restul trupului? Dacă mintea lui Lee era cumva reductibilă la un organ din corp, cu siguranţă că şi a mea era. Asta însemna că eu – fiinţa mea conştientă – voi exista doar o perioadă relativ scurtă.

Apoi voi dispărea.

Întreaga mea carieră ştiinţifică m-am tot gândit la această problemă. Am vrut să înţeleg ce i s-a întâmplat fratelui meu şi ce ni se va întâmpla în timp tuturor. Aveam nevoie să înţeleg, în termeni biologici, ce înseamnă existenţa noastră ca fiinţe care trăiesc ceva. Pe scurt, să înţeleg conştiinţa. Din acest motiv mam specializat în domeniul neuroştiinţelor.

Chiar şi astăzi cred că nu aş fi putut urma un drum mai direct spre răspunsurile pe care le căutam.

S-ar putea ca natura conştiinţei să fie cel mai dificil subiect al ştiinţei. El contează pentru că fiecare *este* conştiinţa lui, dar e controversat din cauza a două enigme ce îi chinuie pe gânditori de secole. Prima este întrebarea care e legătura dintre minte şi corp – sau, pentru cei cu o înclinaţie materialistă (adică aproape toţi specialiştii în neuroştiinţe), cum creează creierul mintea. Aceasta este numită „problema minte/corp”. Cum produce creierul fizic experienţa fenomenală? Şi încă o întrebare la fel de derutantă: cum reuşeşte chestia aia nonfizică numită conştiinţă să controleze corpul fizic?

Filosofii încadrează această problemă în domeniul pe care ei îl numesc „metafizică”, adică un fel de a spune că ei nu cred că poate fi rezolvată din punct de vedere ştiinţific. De ce nu? Deoarece ştiinţa se bazează pe metode empirice, iar „empiric” implică „derivat din dovezi senzoriale”. Mintea nu este accesibilă observaţiei senzoriale. Nu poate fi văzută sau atinsă, este invizibilă şi intangibilă, este o temă, nu un obiect.

A doua enigmă se referă la ce putem şti despre minte din exteriorul ei – la drept vorbind, cum ne putem da seama că mintea există. Această enigmă este numită „problema altor minţi”, adică, dacă minţile sunt subiective, atunci nu o poţi observa decât pe a ta. Şi atunci cum putem să ne dăm seama dacă alţi oameni (sau alte fiinţe sau maşini) au minte, darămite să identificăm legi obiective care guvernează modul în care funcţionează în general minţile?

În ultimul secol, aceste întrebări au primit trei răspunsuri ştiinţifice importante. Ştiinţa se bazează pe experimente. Un lucru în avantajul nostru este că metoda experimentală nu aspiră la adevăruri supreme, ci mai degrabă la ceea ce poate fi descris drept cele mai bune presupuneri. Pornind de la observaţii, facem conjecturi în legătură cu explicaţiile plauzibile ale fenomenelor observate. Cu alte cuvinte, formulăm ipoteze. Apoi generăm *predicţii* pornind de la ipoteze. Acestea iau forma: „Dacă ipoteza X este corectă, atunci, când fac Z, ar trebui să se întâmple Y” (în care există o şansă rezonabilă ca Y să nu se întâmple dacă se pleacă de la alte ipoteze). Acesta este experimentul. Dacă Y nu se întâmplă, atunci se deduce că X este falsă şi este revizuită în conformitate cu observaţiile noi. Apoi, procesul experimental o ia de la capăt, până când dă naştere unor predicţii falsificabile care sunt confirmate. În acest moment, considerăm că ipoteza este *provizoriu* adevărată, până când şi dacă nu o contrazic alte observaţii. În acest fel, nu ne aşteptăm să ajungem la certitudine în ştiinţă, ci aspirăm doar la mai puţină incertitudine[1].

Începând din prima jumătate a secolului XX, o şcoală de psihologie numită „behaviorism” a început să aplice minţii în mod sistematic metoda experimentală. Punctul său de plecare era să elimine din câmpul cercetării tot, cu excepţia evenimentelor observabile în mod empiric. Behavioriştii au eliminat toate discuţiile „mentaliste” despre credinţe şi idei, sentimente şi dorinţe şi şi-au limitat domeniul de studiu la răspunsurile vizibile şi tangibile ale subiectului la stimuli obiectivi. Respingeau cu fanatism descrierile subiective a ceea ce se petrece în interior. Tratau mintea ca pe o „cutie neagră”, despre care nu se puteau şti decât inputurile şi outputurile.

De ce au adoptat o poziţie atât de extremă? Bineînţeles, aceasta era parţial o încercare de a ocoli problema altor minţi. Dacă refuzau de la bun început să accepte orice discuţie despre alte minţi, se înţelegea de la sine că teoriile lor nu puteau fi afectate de îndoielile filosofice atât de răspândite în psihologie. De fapt, au exclus psihicul din psihologie.

Acesta poate părea un preţ cam mare. Dar behaviorismul a fost de la început o doctrină revoluţionară. Behavioriştii nu urmăreau puritatea epistemologică ca scop în sine, ci încercau şi să detroneze puterea dominantă în psihologie la vremea respectivă. Psihanaliştii freudieni dominaseră ştiinţa minţii de la începutul secolului. Examinând îndeaproape caracteristicile neobişnuite ale mărturiilor introspective, Sigmund Freud căutase să dezvolte un model al minţii analizate, cum ar veni, dinspre interior

spre exterior. Ideile rezultate au stabilit agenda cu privire la tratament şi cercetare pentru o jumătate de secol, generând instituţii, experţi acreditaţi şi un nucleu de intelectuali deschizători de drumuri remarcabili. Totuşi, behavioriştii considerau că toate teoriile lui Freud nu erau decât nişte fantezii brodate pe bazele fragile ale subiectivităţii. Freud se aruncase cu capul înainte în problema altor minţi şi târâse după el şi restul psihologiei. Behavioriştii şi-au propus să o scoată la liman.

În pofida austerităţii programului lor, ei chiar au reuşit să deducă relaţii cauzale dintre anumite tipuri de răspunsuri şi stimuli mintali. Şi nu numai: au reuşit să manipuleze inputurile astfel încât să obţină schimbări previzibile la nivelul outputurilor. Astfel, au descoperit unele dintre legile fundamentale ale învăţării. De pildă, când declanşatorul unui comportament involuntar este însoţit în mod repetat de un stimul artificial, atunci stimulul artificial va ajunge să declanşeze acelaşi răspuns involuntar ca stimulul natural. Aşadar, dacă imaginea hranei este însoţită în mod repetat de sunetul unui clopoţel (la animalele care salivează în mod natural atunci când văd hrană, precum câinii), atunci sunetul clopoţelului va ajunge să declanşeze salivarea şi singur. Aceasta se numeşte „condiţionare clasică”. De asemenea, dacă un comportament voluntar este însoţit în mod repetat de recompense, comportamentul respectiv va apărea mai frecvent, iar dacă acelaşi comportament este însoţit de pedepse, se va manifesta mai rar. Deci, dacă un câine ce sare pe oaspeţi primeşte îmbrăţişări, va sări şi mai mult pe ei; dacă primeşte o

lovitură, va sări mai puţin. Acest lucru se numeşte „condiţionare operantă” – sau legea efectului.

Asemenea descoperiri nu au fost realizări mărunte. Ele au demonstrat că mintea se supune unor legi naturale, ca orice altceva. Dar mintea înseamnă mult mai mult decât învăţarea şi chiar învăţarea este influenţată şi de alţi factori în afară de stimulii externi. Imaginează-ţi că îţi spui: „După ce citesc această pagină, îmi fac un ceai”. Acest tip de gândire îţi influenţează comportamentul tot timpul. Behavioriştii nu considerau însă astfel de rapoarte introspective date ştiinţifice acceptabile, deoarece gândurile nu sunt observabile din exterior. În consecinţă, nu aveau de unde să ştie ce te-a determinat să-ţi faci un ceai.

Marele neurolog Jean-Martin Charcot a spus la un moment dat: „Teoria e bună, dar nu împiedică lucrurile să existe”[2]. Întrucât evenimentele mintale interne există în mod clar şi influenţează cauzal comportamentul, abordarea behavioristă a fost eclipsată treptat în a doua jumătate a secolului XX de alta. Aceasta a fost numită psihologie „cognitivă”, termen ce include procesele psihice interioare, ca să spunem aşa.

Stimulentul din spatele revoluţiei cognitive a fost apariţia calculatoarelor. Behavioriştii considerau că funcţionarea minţii constituie o „cutie neagră” impenetrabilă, axându-se, în schimb, pe inputuri şi outputuri. Dar calculatoarele nu sunt impenetrabile. Ne-ar fi fost imposibil să le inventăm fără să le înţelegem pe deplin funcţionarea internă. De aceea, pentru că tratau mintea ca pe un calculator, psihologii s-au simţit încurajaţi să formuleze modele ale *prelucrării informaţiilor* care se petrece în

interiorul ei. Modelele lor au fost apoi testate cu ajutorul unor simulări artificiale ale proceselor psihice, în combinaţie cu experimente behavioriste.

Ce este prelucrarea informaţiilor? Voi spune mai multe despre ea mai jos, dar cel mai interesant e că poate fi implementată cu tipuri foarte diferite de echipamente fizice. Astfel, reuşim să vedem natura fizică a minţii într-o lumină nouă. Ni se sugerează că mintea (concepută ca prelucrare a informaţiilor) este o *funcţie,* şi nu o structură. Din această perspectivă, funcţiile de „software” ale minţii sunt implementate de structurile „hardware” ale creierului, dar aceleaşi funcţii pot fi implementate la fel de bine şi de alte substraturi, precum calculatoarele. Astfel, atât creierii, cât şi calculatoarele îndeplinesc funcţii de *memorie* (codează şi stochează informaţii) şi de *percepţie* (clasifică tipare ale informaţiilor de input prin compararea lor cu informaţiile stocate), precum şi funcţii de *execuţie* (execută decizii referitoare la ce este de făcut ca reacţie la asemenea informaţii).

Aceasta este puterea a ceea ce a ajuns să se numească abordarea „funcţionalistă”, dar este şi punctul ei slab. Dacă aceleaşi funcţii pot fi îndeplinite şi de calculatoare, despre care presupunem că nu sunt fiinţe conştiente, atunci se justifică să reducem mintea la simpla prelucrare a unor informaţii? Până şi telefonul tău are funcţii de memorie, de percepţie şi de execuţie.

Al treilea răspuns ştiinţific important la metafizica minte/corp s-a dezvoltat în tandem cu psihologia cognitivă, dar, până la sfârşitul secolului trecut, a ajuns să o eclipseze. Mă refer la o abordare numită în general

„neuroştiinţă cognitivă”. Aceasta se axează pe partea de hardware a minţii şi a apărut odată cu dezvoltarea a numeroase tehnici psihologice care ne permit să observăm şi să măsurăm direct dinamica creierului viu.

Pe vremea behaviorismului, neurofiziologii erau limitaţi la o singură asemenea tehnică: puteau să înregistreze activitatea electrică a creierului de la suprafaţa exterioară a scalpului cu ajutorul unei electroencefalograme (EEG). Astăzi, avem mult mai multe instrumente la dispoziţie, precum imagistica prin rezonanţă magnetică funcţională (RMNf), pentru a măsura frecvenţa activităţii hemodinamice în diferite părţi ale creierului în timp ce execută operaţii mintale specifice, şi tomografia cu emisie de pozitroni (PET CT), cu care putem măsura activitatea metabolică diferenţială pentru fiecare sistem de neurotransmiţători. Astfel avem posibilitatea să identificăm cu precizie ce procese cerebrale ne generează diferitele stări psihice. De asemenea, avem posibilitatea să vizualizăm conectivitatea funcţional-anatomică detaliată între diferitele regiuni cerebrale folosind tractografia. Iar cu ajutorul optogeneticii putem să vedem şi să activăm circuitele de neuroni ce alcătuiesc urmele unor amintiri individuale, pe măsură ce luminează în timpul unor sarcini cognitive.

Aceste tehnici fac vizibilă clar funcţionarea internă a organului minţii – astfel concretizând cele mai hazardate visuri empiriste ale behavioriştilor, fără a limita sfera psihologiei la stimuli şi răspunsuri.

Stadiul neuropsihologiei în anii 1980, când am intrat

eu în acest domeniu, explică de ce behavioriştii au făcut o tranziţie atât de uşoară de la teoria învăţării la neuroştiinţa cognitivă. Neuropsihologia de la acea vreme putea fi numită la fel de bine neurobehaviorism. Cu cât ni se preda despre funcţii precum memoria pe termen scurt, despre care se spunea că este un „amortizor” pentru reţinerea amintirilor în conştiinţă, cu atât realizam că profesorii mei vorbeau despre altceva decât ceea ce urmăream eu să aflu. La cursuri ni se vorbea despre instrumentele funcţionale folosite de minte, şi nu despre minte în sine. Nu ştiam ce să mai înţeleg.

Neurologul Oliver Sacks, în cartea sa *Un picior de sprijin* (1984), a descris sugestiv situaţia în care mă aflam:

> Neuropsihologia, ca şi neurologia clasică, urmăreşte să fie cu totul obiectivă, iar marea sa forţă, progresele sale tocmai de aici vin. Dar o fiinţă vie, mai ales o fiinţă umană, este în primul rând activă – subiect, nu obiect. Tocmai subiectul, „eul” viu, este exclus. Neuropsihologia e admirabilă, dar exclude psihicul – exclude „eul” viu, activ, ce are experienţe[3].

Afirmaţia „Neuropsihologia e admirabilă, dar exclude psihicul” mi-a surprins perfect dezamăgirea. După ce am citit-o, am iniţiat o corespondenţă cu Oliver Sacks care a continuat până la moartea sa, în 2015. M-a atras la el faptul că lua atât de în serios relatările subiective ale pacienţilor lui. Acest lucru era deja evident în cartea sa din 1970, *Migraine,* şi chiar mai mult în

extraordinarul său volum *Revenirea la viaţă* (1973). Cea de-a doua carte consemna cu detalii de mare fineţe parcursul clinic al unui grup de pacienţi cronici cu achinezie, mutism şi encefalită letargică. Această afecţiune era numită şi „boala somnului”, deşi pacienţii nu dormeau propriu-zis, ci doar nu manifestau iniţiativă sau pulsiune spontană. Sacks „i-a trezit” administrându-le levodopă, o substanţă care creşte cantitatea de dopamină. Totuşi, după ce au revenit la o intenţionalitate activă, au devenit rapid excesiv de energici, maniacali şi, în cele din urmă, psihotici. La scurt timp după ce am citit *Un picior de sprijin,* care descrie experienţa subiectivă a unei leziuni a sistemului nervos trăită chiar de Sacks, acesta a publicat *Omul care îşi confunda soţia cu o pălărie* (1985) – o serie de studii de caz ce furnizează introspecţii revelatoare legate de diverse afecţiuni neuropsihologice din perspectiva unor pacienţi cu antecedente neurologice. Astfel Sacks a dobândit o faimă de durată.

Aceste cărţi nu prea semănau cu manualele mele de neuropsihologie, ce disecau funcţiile mintale în modul în care am proceda în cazul funcţiilor oricărui organ. De exemplu, am aflat că limbajul este produs în aria Broca din lobul frontal stâng, că înţelegerea vorbirii are loc în aria Wernicke, la câţiva centimetri mai în spate, în lobul temporal, şi că abilitatea de a repeta ce ţi se spune este mediată de fasciculul arcuat, un tract fibros ce conectează aceste două arii. De asemenea, am aflat că amintirile sunt codificate de hipocamp, stocate în neocortex şi accesate prin mecanisme frontale-limbice.

Chiar nu este creierul diferit de stomac şi de

plămâni? Aspectul evident care îl individualizează e faptul că există „ceva cum e” *să fii* un creier. Acest lucru nu se aplică şi altor părţi ale corpului. Senzaţiile pe care le localizăm în alte organe nu sunt simţite de acestea, ci impulsurile nervoase care pornesc din ele sunt simţite numai când ajung la creier. Cu siguranţă exista un motiv pentru care ţesutul cerebral are această proprietate foarte specifică – capacitatea de a percepe, a simţi şi a gândi lucruri. Această proprietate pare că face ceva. Şi, dacă face – dacă experienţele subiective au efecte cauzale asupra comportamentului, aşa cum pare când te hotărăşti spontan să-ţi faci un ceai –, atunci am fi într-o mare eroare dacă am omite-o din descrierile ştiinţifice. Totuşi, exact asta se întâmpla în anii 1980. Profesorii nu ne-au spus nimic niciodată despre cum e să înţelegi vorbirea sau să accesezi o amintire, ca să nu mai vorbim de motivul pentru care simţim ce simţim.

Cei care ţineau seama de perspectiva subiectivă nu erau luaţi în serios de specialiştii în neuroştiinţe. Nu ştiu câţi oameni cunosc faptul că lucrările lui Sacks erau în mare parte ridiculizate de colegii lui. Unul dintre comentatori chiar l-a numit „Omul care şi-a confundat pacienţii cu o carieră literară”. Toate acestea l-au afectat în mare măsură. Cum să descrii viaţa interioară a fiinţelor umane fără să le spui poveştile? După cum se lamentase Freud cu un secol înainte în privinţa rapoartelor sale clinice:

Continuă să mi se pară ciudat faptul că uneori istoricul bolii scris de mine poate fi citit ca o nuvelă şi

că aceasta este lipsită, ca să spunem aşa, de pecetea serioasă a travaliului ştiinţific. Trebuie să mă consolez cu faptul că, evident, pentru acest rezultat poate fi făcută răspunzătoare mai degrabă natura obiectului decât preferinţa mea*[1] 4.

Sacks a fost încântat când i-am trimis acest citat[5]. În ce mă priveşte, când am citit aceste rânduri pentru prima dată, mi-am dat seama că nu sunt singurul care a intrat în domeniul neuropsihologiei cu speranţa că îmi va permite să învăţ cum generează creierul subiectivitatea. Ajungi repede să nu mai crezi acest lucru. Eşti avertizat să nu mai urmăreşti asemenea probleme de nerezolvat – ele sunt „nocive pentru cariera ta”. Şi, astfel, majoritatea celor care studiază neuroştiinţele uită treptat de ce au intrat în acest domeniu şi ajung să se identifice cu dogma cognitivismului, care abordează creierul ca şi cum nu ar fi deloc diferit de un telefon mobil.

Singurul aspect al conştiinţei ce era într-adevăr o temă ştiinţifică respectabilă în anii 1980 era mecanismul cerebral al stării de veghe versus starea de somn. Cu alte cuvinte, „nivelul” de conştiinţă era un subiect respectabil, dar nu şi „conţinutul” ei. Aşadar, m-am hotărât să-mi axez activitatea de cercetare doctorală pe un aspect al somnului. În particular, am ales să studiez aspectul subiectiv al somnului, şi anume mecanismele cerebrale ale visului. Până la urmă, visele nu sunt decât o imixtiune paradoxală a conştiinţei („starea de veghe”) în starea de somn. În mod uimitor, exista o lacună semnificativă în literatura de specialitate pe acest subiect: nimeni nu descrisese sistematic cum

anume degradarea unor părţi ale creierului afectează visul. Aşa că mi-am propus să o fac eu.

Visul este dificil de studiat tocmai din cauza naturii sale subiective. Fenomenele mintale în general nu pot fi observate decât introspectiv, de un singur observator, apoi raportate altora indirect, prin cuvinte. Dar visele sunt şi mai problematice: nu pot fi raportate decât retrospectiv, după ce visul se încheie şi cel care visează se trezeşte. Toată lumea ştie că nu ne putem baza pe amintirile noastre despre vise. Ce fel de „date” sunt acestea[6]? Acesta este şi motivul pentru care, începând de la mijlocul secolului XX, visele au constituit un front semnificativ în tranziţia de la behaviorism la ceea ce avea apoi să devină neuroştiinţa cognitivă.

Electroencefalograma a fost aplicată pentru prima dată în studiul somnului la începutul anilor 1950 de doi neurofiziologi, Eugene Aserinsky şi Nathaniel Kleitman. Ei au plecat de la ipoteza că nivelul activităţii cerebrale scade atunci când adormim şi creşte când ne trezim, prezicând de aici că amplitudinea undelor noastre cerebrale (unul dintre lucrurile măsurate prin electroencefalografie) creşte, iar frecvenţa lor (celălalt lucru pe care îl măsoară) scade pe măsură ce adormim, precum şi că se întâmplă invers când ne trezim (vezi figura 10, p. 136).

Când creierul intră în ceea ce numim acum somn „cu unde lente”, înţelegem exact ceea ce au prevăzut Aserinsky şi Kleitman. Ipoteza lor s-a confirmat. Surpriza constă în ce se întâmplă în continuare: în decurs de aproximativ 90 de minute după ce adormim

(apoi la fiecare circa 90 de minute, în cicluri regulate), undele cerebrale accelerează din nou, aproape atingând nivelul din starea de veghe, chiar dacă persoana de la care se obţine înregistrarea rămâne în starea de somn[7]. Aserinsky şi Kleitman au numit aceste etape ciudate ale activării creierului „somn paradoxal” – paradoxul fiind că creierul este stimulat din punct de vedere fiziologic, deşi doarme profund.

Se mai întâmplă tot felul de alte lucruri în această stare aparte. Ochii se mişcă rapid (motivul pentru care somnul paradoxal a fost ulterior denumit somn „cu mişcare rapidă a ochilor” sau somn REM), dar corpul de la gât în jos este paralizat temporar. Au loc şi modificări involuntare spectaculoase, precum controlul redus al temperaturii bazale şi mărirea organelor genitale, ducând la erecţii vizibile la bărbaţi. Nu reuşesc să înţeleg cum de nu a observat ştiinţa toate acestea până în 1953.

Pe baza acestor observaţii, Aserinsky şi Kleitman au formulat o altă ipoteză, despre care nu se poate spune că nu era fondată: somnul REM este baza fiziologică a stării psihologice numite vis. În consecinţă, au prezis că, dacă suntem treziţi din somnul REM, vom putea povesti vise, dar nu şi dacă suntem treziţi din somnul cu unde lente (non-REM). Împreună cu cercetătorul care a avut ghinionul să se numească William Dement, au testat această predicţie şi au confirmat-o: aproximativ 80% dintre trezirile din somnul REM şi mai puţin de 10% din trezirile din somnul non-REM au produs evocarea unor vise. Din acel moment, somnul REM a fost considerat sinonim cu visul[8]. Ce veste excelentă! Domeniul nu mai

trebuia să se preocupe de vis, deoarece acum avea un indicator obiectiv al acestuia, care le permitea specialiştilor în neuroştiinţe să-şi desfăşoare activitatea ca la carte, fără a mai fi nevoiţi să se confrunte cu complicaţiile metodologice introduse de evocările verbale retrospective, cu un singur martor ale unor experienţe subiective efemere.

Mai exista un motiv de bucurie că scăpaserăm de vise. Acesta consta în rolul stânjenitor pe care îl jucaseră în consacrarea psihanalizei. Spre deosebire de răspunsurile clasice la metafizica minte/corp ce caracterizau ştiinţa minţii în a doua jumătate a secolului XX, psihanaliştii nu aveau reţineri în a trata relatările introspective drept date. De fapt, relatările obţinute prin „asociere liberă” (testarea nestructurată a fluxului conştiinţei) constituiau principalele date ale studiilor psihanalitice. Folosind această metodă, Sigmund Freud a ajuns la concluzia că, în pofida aparenţei absurde a unor experienţe onirice „manifeste”, conţinutul lor „latent” (povestea de bază, dedusă de Freud din asocierile libere ale persoanei care visa) dezvăluia o funcţie psihologică coerentă. Această funcţie era *împlinirea dorinţelor.*

Conform lui Freud, visul este ceea ce se întâmplă atunci când nevoile biologice care generează comportamentul din starea de veghe sunt dezinhibate în timpul somnului. Visele sunt încercări de satisfacere a acestor nevoi, care continuă să ne solicite chiar şi în timpul somnului. Totuşi, visele fac acest lucru într-un mod halucinatoriu, permiţându-ne astfel să rămânem adormiţi (în loc să ne trezim ca să ne satisfacem cu

adevărat pulsiunile). Întrucât halucinaţiile sunt o caracteristică esenţială a tulburărilor mintale, Freud a folosit această teorie în lucrarea sa de referinţă *Interpretarea viselor* (1900) ca să descrie un model general al modului de funcţionare a minţii în ansamblul ei, în condiţii de sănătate şi de boală.

După cum a spus Freud, „psihanaliza se bazează pe analiza visului”[9]. Dar visele, după cum am văzut, sunt incredibil de dificil de studiat în mod empiric, aşa că behavioriştii le-au eliminat din ştiinţă. În plus, edificiul teoretic construit de Freud pe vise nu era mai bun decât fundamentele sale. Marele filosof al ştiinţei Karl Popper a declarat teoria psihanalitică drept „pseudoştiinţifică”, din cauză că nu genera predicţii falsificabile din punct de vedere experimental[10]. Cum falsifici ipoteza dedusă de Freud, conform căreia visele exprimă dorinţe latente? Dacă nu e neapărat ca dorinţele să apară în visul manifest (relatat), atunci orice vis poate fi „interpretat” pentru a se potrivi cerinţelor teoriei. Aşadar, deloc surprinzător, când descoperirea somnului REM le-a permis specialiştilor în neuroştiinţe să treacă de la materia efemeră a evocărilor viselor la corelativele lor fiziologice concrete, visele le-au scăpat printre degete.

Descoperirea somnului REM în anii 1950 a declanşat o cursă pentru identificarea bazei sale neurologice, întrucât funcţionarea somnului REM putea dezvălui mecanismul *obiectiv* al viselor, a cărui elucidare putea să plaseze psihiatria vremii pe o poziţie ştiinţifică mai respectabilă. (Această cercetare a fost simplificată datorită faptului că somnul REM apare la toate

mamiferele). Cursa a fost câştigată de Michel Jouvet în 1965. Într-o serie de experimente chirurgicale pe pisici, Jouvet a demonstrat că somnul REM era generat nu de creierul anterior (ce include cortexul, partea exterioară a creierului care este atât de impresionant de mare la oameni şi, parţial din acest motiv, este considerată organul minţii), ci de trunchiul cerebral, o structură presupusă a fi mult mai modestă, de o origine evolutivă extrem de veche[11]. Jouvet a ajuns la această concluzie observând că secţiunile progresive de la nivelul creierului, începând din partea superioară şi continuând în jos, produceau pierderea somnului REM numai atunci când secţionarea ajungea la nivelul unei structuri „inferioare” a trunchiului cerebral numită punte (vezi figura 1)[12].

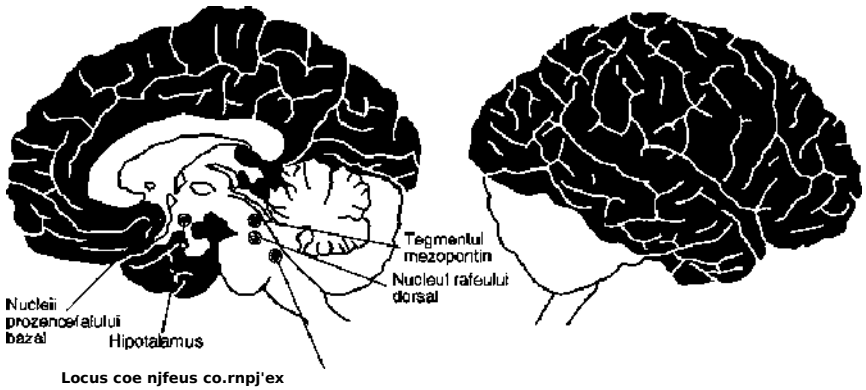


Figura 1. Imaginea din stânga este o reprezentare mediană a
creierului (secţionat prin mijloc), iar cea din dreapta este o
reprezentare laterală (văzut dintr-o parte). Figura arată cortexul (negru) şi trunchiul cerebral (alb). Sunt indicaţi numai

nucleii trunchiului cerebral consideraţi importanţi pentru
controlul somnului REM, şi anume tegmentul mezopontin,
nucleul rafeului dorsal şi *locus coeruleus complex.* Este reprezentată şi localizarea nucleilor prozencefalului bazal (sub
cortex) şi a hipotalamusului, a cărei relevanţă se va vedea mai
jos.

Sarcina de a se ocupa de detalii i-a revenit lui Allan Hobson, un student al lui Jouvet. Hobson a identificat exact grupările de neuroni pontini care generau somnul REM şi, prin urmare, visele. Până pe la mijlocul anilor 1970 a devenit evident că întregul ciclu somn/stare de veghe – inclusiv toate fenomenele somnului REM enumerate mai sus, precum şi cele din diferitele etape ale somnului non-REM – era orchestrat de un număr mic de nuclei din trunchiul cerebral care interacţionau între ei[13]. Cei care controlează somnul REM semănau cu un întrerupător simplu de tip pornit/oprit. Neuronii care *declanşează* somnul REM se află în tegmentul mezopontin (vezi figura 1). Aceştia secretă în tot creierul anterior o substanţă neurochimică numită acetilcolină. Acetilcolina provoacă activarea: creşte „nivelul” conştiinţei (de exemplu, este stimulată de nicotină, ceea ce te ajută să te concentrezi). Neuronii din trunchiul cerebral care *opresc* somnul REM se află mai în profunzimea punţii, în rafeul dorsal şi în *locus*

coeruleus complex (din nou, vezi figura 1). Ei eliberează serotonină şi, respectiv, noradrenalină. La fel ca acetilcolina, aceşti compuşi neurochimici modulează diferite aspecte ale nivelului de conştiinţă.

Combinând aceste concluzii cu faptul că somnul REM se declanşează şi se opreşte automat, cam la fiecare 90 de minute, ca un ceas, Hobson a conchis inevitabil: „Principala forţă ce declanşează visele nu este de natură psihologică, ci fiziologică, deoarece momentul apariţiei şi durata somnului cu vise sunt destul de constante, sugerând o geneză preprogramată, determinată neuronal”[14].

Deoarece somnul REM apare din trunchiul cerebral colinergic, o parte străveche şi inferioară a creierului departe de maiestuosul cortex unde se presupune că are loc toată acţiunea psihologiei umane, Hobson a adăugat că visul nu poate fi stimulat de dorinţe; e „neutru din punctul de vedere al stimulării”[15]. Prin urmare, conform lui Hobson, opinia lui Freud că visele vin din dorinţe latente este probabil complet greşită. Semnificaţia pe care Freud a văzut-o în vise nu le era mai intrinsecă decât le este petelor de cerneală. Aceasta era proiectată asupra lor, neexistând în visul în sine. Din punct de vedere ştiinţific, interpretarea viselor nu era cu nimic mai bună decât ghicitul în frunze de ceai.

Deoarece întreaga psihanaliză era fundamentată pe metoda folosită de Freud pentru a studia visele, toate teoriile derivate astfel de acesta puteau fi respinse. Ca urmare a demontării de către Hobson a ideii că visele ar avea vreo valoare, psihiatria putea, în sfârşit, să se îndepărteze de fundamentarea ei istorică pe relatările

introspective şi să se bazeze de-acum pe metodele obiective ale neuroştiinţelor (mai ales neurochimia) pentru cercetare şi tratament. În consecinţă, dacă în anii 1950 era aproape imposibil să devii profesor titular de psihiatrie la o universitate americană importantă fără să fii şi psihanalist, astăzi e invers: este aproape imposibil să devii profesor de psihiatrie dacă *eşti* psihanalist.

Nimic din toate acestea nu m-a surprins în mod deosebit la vremea respectivă. Problema din centrul cercetărilor mele doctorale părea destul de clară şi deloc implicată în luptele pentru moştenirea freudismului şi a behaviorismului. Eu nu voiam să ştiu decât cum este afectată experienţa reală a visului de leziunile la nivelul diferitelor părţi ale creierului anterior şi ale cortexului său. Până la urmă, dacă acţiunea avea loc în prozencefal, din punct de vedere psihologic, cu siguranţă că trebuia să facă *ceva* în vise.

Departamentul de neurochirurgie din cadrul University of Witwatersrand avea secţii în două spitale universitare – Baragwanath Hospital şi Johannesburg General Hospital. Baragwanath era un fost spital militar mare, în complexul urban „noneuropean” Soweto. Având în vedere că toate acestea
se întâmplau la apogeul apartheidului, era un ocean de suferinţă umană. În schimb, Johannesburg General Hospital, rezervat „europenilor”, era un spital universitar modern, un monument închinat inegalităţii rasiale. De asemenea, departamentul de neurochirurgie avea paturi şi în secţia de recuperare după leziuni

cerebrale şi spinale din cadrul Edenvale General Hospital, care funcţiona într-o veche clădire colonială dintr-o suburbie a Johannesburgului. Începând din 1985, am lucrat în toate aceste trei locuri, examinând sute de pacienţi anual. I-am inclus pe 361 dintre ei în studiile mele doctorale, care s-au desfăşurat în următorii cinci ani.

După ce am învăţat să folosesc tehnologia electroencefalografiei şi altele conexe şi să recunosc undele cerebrale caracteristice asociate cu diferitele etape ale somnului, am reuşit să-i trezesc pe subiecţi în timpul somnului REM, când era cea mai mare probabilitate să viseze. De asemenea, am chestionat pacienţi cu antecedente neurologice aflaţi la pat despre modificări ale viselor lor, apoi i-am monitorizat zile, săptămâni şi luni la rând. Astfel am investigat dacă conţinutul viselor era afectat sistematic de leziunile localizate în diferite arii cerebrale. În pofida reputaţiei îndoielnice a relatărilor viselor, am presupus că, dacă pacienţii cu leziuni la nivelul aceleiaşi arii cerebrale reclamau aceeaşi modificare a conţinutului viselor, aveam toate motivele să-i cred. Această metodă se numeşte „corelare clinico-anatomică”: prin examinarea clinică a abilităţilor psihologice ale pacienţilor, observi cum se modifică o funcţie cerebrală ca urmare a unei leziuni la nivelul unei anumite părţi a creierului; apoi corelezi această modificare cu locul leziunii, astfel descoperind indicii cu privire la funcţia structurii cerebrale lezate, ceea ce duce la ipoteze testabile. Această metodă se aplica sistematic de multe decenii în cazul tuturor funcţiilor cognitive principale, precum

percepţia, memoria şi limbajul, dar încă nu fusese aplicată în cazul visului.

La început, am fost puţin stânjenit să vorbesc cu oameni atât de grav bolnavi despre visele lor. Pe mulţi dintre ei îi aştepta sau tocmai trecuseră printr-o intervenţie chirurgicală la nivelul creierului ce le putea pune viaţa în pericol şi, în asemenea condiţii, mă temeam să nu îmi considere întrebările frivole. Dar pacienţii mei erau surprinzător de dispuşi să descrie modificările din viaţa lor psihică generate de afecţiunile neurologice.

Până să-mi încep cercetările, fuseseră publicate mai multe studii de caz în care efectul observat în experimentele pe animale apărea şi la oameni, şi anume că somnul REM nu mai survenea dacă existau leziuni la nivelul tegmentului mezopontin (vezi figura 1). M-a surprins însă că nu se gândise nimeni să pună întrebări despre modificările survenite în *visele* pacienţilor. Este un exemplu cât se poate de clar al ideilor preconcepute din neuroştiinţe faţă de datele subiective[16].

În cercetarea mea, m-am aşteptat să descopăr ceea ce era evident: că pacienţii cu leziuni la nivelul cortexului vizual au vise nonvizuale, că pacienţii cu leziuni la nivelul cortexului responsabil de limbaj au vise nonverbale, că pacienţii cu leziuni la nivelul cortexului somatosenzorial şi motor au vise hemiplegice şi aşa mai departe. Acestea sunt ideile de bază în ce priveşte corelaţia creier-comportament. Acesta era golul pe care voiam să-l umplu şi am făcut-o bucuros[17].

Totuşi, spre uimirea mea, pe lângă toate chestiunile evidente pe care le-am observat, am descoperit şi că

pacienţii cu leziuni la nivelul ariei cerebrale care generează somnul REM *încă aveau vise.* Mai mult, pacienţii care *nu* mai visau suferiseră leziuni în arii cerebrale complet diferite. Prin urmare, visul şi somnul REM erau ceea ce numim fenomene „dublu disociabile”—. Se corelau reciproc (adică de obicei surveneau în acelaşi timp), dar nu erau acelaşi lucru[19].

Timp de aproape 50 de ani, în domeniul ştiinţei somnului, cei care studiau creierul au confundat corelaţia cu identitatea. De îndată ce au stabilit că visul însoţeşte somnul REM, s-au grăbit să tragă concluzia că sunt unul şi acelaşi lucru – apoi s-au debarasat de partea subiectivă problematică a corelaţiei. Din acel moment, cu foarte puţine excepţii, au studiat exclusiv somnul REM, în principal în cadrul unor experimente pe animale, care nu pot furniza relatări introspective. Eroarea a ieşit la suprafaţă abia atunci când am început să studiez din perspectiva neuroştiinţelor *experienţa* viselor la pacienţii cu antecendente neurologice.

La începutul anilor 1990, când am semnalat pentru prima dată că visele dispar odată cu leziunile la nivelul altei arii cerebrale decât cea care generează somnul REM, am insistat să subliniez faptul că aria critică nu se află în trunchiul cerebral[20]. Am procedat astfel deoarece voiam să evidenţiez natura mintală a viselor, şi ştiam toţi că sediul funcţiilor mintale este cortexul.

De fapt, am identificat două regiuni în care leziunile duc la pierderea viselor cu păstrarea somnului REM. Prima se află în cortex, în lobulul parietal inferior (vezi figura 2). Această descoperire nu a fost surprinzătoare, întrucât lobul parietal este important pentru memoria

pe termen scurt. Dacă un pacient nu îşi poate menţine conţinutul memoriei în amortizorul conştiinţei, cum să viseze? Mult mai interesantă a fost a doua regiune cerebrală, şi anume substanţa albă din cadranul ventromedial al lobilor frontali, care conectează cortexul frontal cu diferite structuri subcorticale. Această descoperire a fost cu totul neaşteptată. Nimic din funcţiile acestei regiuni cerebrale nu are o conexiune evidentă cu experienţa manifestă a visului, dar cu siguranţă contribuie cu ceva esenţial la acest proces, întrucât leziunile la nivelul acestei regiuni au cauzat în mod sigur dispariţia totală a viselor.

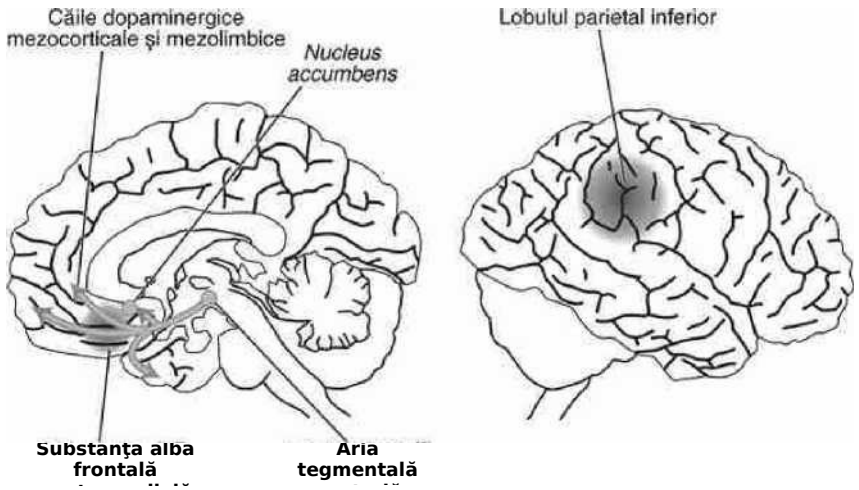


Figura 2. Cele două regiuni cu leziuni care au dus la oprirea
viselor sunt reprezentate cu o umbră în această figură: substanţa albă ventromedială din lobul frontal (stânga) şi
lobulul parietal inferior al cortexului (dreapta). Sunt reprezentate şi aria tegmentală ventrală a trunchiului cerebral

şi principalele căi fibroase ce pornesc de acolo, şi anume căile
dopaminergice mezocorticale şi mezolimbice. Notă: locul
leziunii la nivelul lobului frontal ventromedial implică aceste
căi subcorticale, care trec pe sub cortex, şi nu prin el. O destinaţie importantă a acestor căi este *nucleus accumbens,* de
asemenea reprezentat.

Spun „în mod sigur”, deşi am raportat doar nouă cazuri de pierdere a viselor în rândul pacienţilor mei cu leziuni ale lobului frontal (şi 44 de cazuri pentru leziunile parietale). Asemenea leziuni sunt extrem de rare în practica clinică obişnuită. Totuşi, corelaţia este sigură. În prima jumătate a secolului XX, substanţa albă frontală ventromedială a fost vizată chirurgical în mii de cazuri printr-o tehnică numită leucotomie prefrontală modificată. Psihiatrii acelor zile pline de entuziasm au descoperit că unele afecţiuni mintale grave puteau fi ameliorate prin distrugerea chirurgicală completă a lobilor prefrontali (cunoscută sub denumirea tehnică de lobotomie frontală), dar au observat şi că această procedură radicală are numeroase „efecte secundare”, în termenii lor eufemistici. Prin urmare, au redus amploarea leziunii, încercând să identifice care este cea mai mică parte a lobilor frontali ce poate fi deconectată de restul creierului pentru a se obţine totuşi rezultatele dorite. Procedura modificată de Walter Freeman şi James Watts a fost soluţia. Aceasta presupunea

inserarea unui bisturiu mic cu vibraţii prin orbitele ochilor, care secţiona substanţa albă din cadranul ventromedial al lobilor frontali (leucotomie prefrontală), exact în locul leziunilor celor nouă pacienţi ai mei.

Aşadar, m-am întors la vechea literatură despre psihochirurgie ca să văd dacă îmi confirmă ce am observat la cazurile mele[21]. Aveam motive să sper că doctorii care au examinat pacienţii cu leucotomie clasică i-au întrebat despre vise după intervenţii – până la urmă, psihiatrii încă luau visele în serios pe atunci. S-a dovedit că am avut dreptate. Aceştia au descoperit că leucotomia prefrontală avea trei efecte psihologice principale. În primul rând, reducea simptomele psihotice pozitive (halucinaţiile şi delirurile). În al doilea rând, reducea stimularea. În al treilea rând, cauza pierderea viselor. De fapt, unul dintre primii cercetători în domeniul psihochirurgiei a sugerat chiar că menţinerea capacităţii de a visa după intervenţie era un semn de prognostic nefavorabil[22].

Această ultimă idee m-a ajutat să intuiesc care dintre numeroasele circuite neuronale din cadranul ventromedial al lobilor frontali era cel mai probabil să fie responsabil pentru pierderea viselor. De asemenea, mi-a furnizat un prim indiciu în ce priveşte motivul pentru care trebuie să căutăm vinovatul în această regiune cerebrală la care nu ne-am aştepta. Ce sunt visele dacă nu halucinaţii şi deliruri? Iată de ce ar fi un prognostic nefavorabil dacă acestea ar persista după leucotomie.

Tratamentul neurochirurgical al halucinaţiilor şi delirurilor nu a fost abandonat din raţiuni etice, ci a

căzut în dizgraţie când a început să se vadă că se pot obţine rezultate terapeutice echivalente cu o rată de morbiditate şi mortalitate mai mică folosind medicamente devenite disponibile în masă prima dată în anii 1950, şi anume „tranchilizantele majore”. Aceste medicamente şi „antipsihoticele” moderne blochează dopamina neurochimică la capetele unui circuit cerebral numit sistemul dopaminergic mezocortical şi mezolimbic (vezi figura 2). Întrucât acest circuit este întrerupt prin leucotomia prefrontală, cum s-a întâmplat la cei nouă pacienţi ai mei cu leziuni survenite natural, am presupus că acesta ar putea fi sistemul ce generează visele.

Alte experimente mi-au confirmat ipoteza. Se stabilise deja că stimularea farmacologică a acestui circuit creştea frecvenţa, durata şi intensitatea viselor, fără efecte proporţionale asupra somnului REM[23]. Era vorba de levodopă, medicamentul folosit de Oliver Sacks pentru a-şi „trezi” pacienţii postencefalie. Neurologii care folosesc stimulenţi ai dopaminei în tratamentul bolii Parkinson ştiu de multă vreme că trebuie să fie atenţi să nu îşi împingă pacienţii spre psihoză, cum a făcut Sacks, iar apariţia unor vise neobişnuit de intense este adesea primul semn al acestui efect secundar[24]. Observaţiile ulterioare cruciale au fost că neuronii ce alcătuiesc acest circuit (ai cărui corpi celulari se găsesc în aria tegmentală ventrală) se descarcă la o rată maximă în timpul somnului cu vise[25] şi, în acelaşi timp, transportă dopamină în cantităţi maxime către ţintele lor din *nucleus accumbens* (vezi figura 2)[26]. Aşadar, astăzi se consideră în general că

visele pot să apară independent de somnul REM şi că circuitul dopaminergic mezocortical- mezolimbic este într-adevăr principalul stimulent al viselor[27].

Leziunile la nivelul căilor colinergice din cadranul ventromedial al lobilor frontali (care apar din nucleii prozencefalului bazal, vezi figura 1) produc efectul opus faţă de ceea ce se întâmplă când există leziuni la nivelul căilor dopaminergice, şi anume *mai multe* vise, nu mai puţine. Hobson afirmase că acetilcolina era generatorul viselor neutru din punctul de vedere al stimulării, dar acelaşi lucru se întâmplă şi când blochezi acetilcolina pe cale farmacologică, ca atunci când apar leziuni la nivelul căilor. Acum se ştie despre medicamentele anticolinergice – blocanţii acetilcolinei – că provoacă vise *în exces*[28]*.* Cu alte cuvinte, blocarea sistemului neuronal despre care Hobson spunea că este responsabil pentru apariţia viselor are efectul opus celui prognozat de teoria lui.

În scurt timp, a fost clar că neuroştiinţele îi datorau scuze lui Freud. Dacă există o parte a creierului care poate fi considerată responsabilă cu „dorinţele”, aceasta este circuitul dopaminergic mezocortical-mezolimbic. Este orice numai stimulent neutru nu. Edmund Rolls (şi mulţi alţii) numeşte acest circuit sistemul de „recompensare” al creierului[29]. Kent Berridge îl numeşte sistemul „dorinţelor”. Jaak Panksepp îl numeşte sistemul CĂUTĂRII – şi atrage atenţia asupra rolului său pentru funcţia de *căutare a hranei*[30]*.* Acesta este circuitul cerebral responsabil cu „comportamentele de explorare şi de căutare cele mai energice pe care le poate manifesta un animal”[31]. Este şi circuitul care

stimulează visele[32].

Hobson nu a fost amuzat. M-a invitat să-mi prezint concluziile în faţa grupului său de cercetare din cadrul departamentului de neurofiziologie de la Harvard. Iniţial, le-a acceptat şi a publicat o recenzie favorabilă a cărţii pe care am scris-o pe acest subiect în 1997, notând că rezultatele mele clinico-anatomice au fost confirmate până la ultimul detaliu de studiile de neuroimagistică ale lui Allen Braun (vezi figura 3, p. 47)[33]. Apoi şi-a dat seama că aceste evoluţii ar putea reabilita o viziune în mare parte freudiană asupra viselor, punct în care mi-a scris că e dispus să-mi sprijine public concluziile cu condiţia să nu pretind că îi dau dreptate lui Freud. Cam asta se poate spune despre presupusa obiectivitate a neuropsihologiei.

Totuşi, descoperirile mele mai aveau un aspect foarte surprinzător. Când am dat peste el pentru prima dată, nu am acordat mare atenţie faptului că neuronii care activează acest circuit sunt localizaţi în *trunchiul cerebral* (la fel ca şi cei din circuitele care generează somnul REM). După cum am spus, am
vrut să evidenţiez natura mintală a viselor. A fost nevoie ca Allen Braun, specialistul în neuroimagistică pe care tocmai l-am menţionat, să-mi atragă atenţia politicos asupra omisiunii mele. În contextul dezacordului ştiinţific dintre mine şi Hobson în privinţa circuitelor cerebrale care activează procesul visului (dopaminergice sau colinergice), Braun a scris:

> Ciudat este că, după ce argumentează că structurile prozencefalului joacă un rol esenţial în sistemul visului, Solms ajunge să sugereze că circuitele

dopaminergice aferente acestor regiuni generează visele – *prin aceasta plasând declanşatorul viselor înapoi în trunchiul cerebral*[34].

Braun a concluzionat: „Mi se pare că aceşti domni încep să ajungă la un consens”[35]. În anii 1990, credeam, la fel ca restul neuropsihologilor, că toată activitatea psihologică avea loc în cortex, aşa că m-am concentrat asupra faptului că tracturile substanţei albe care mă interesau erau în lobii frontali, unde se localizau leziunile în cele nouă cazuri ale mele. Dar toţi nucleii trunchiului cerebral au axoni lungi orientaţi în sus, spre prozencefal (vezi figura 2). Corpii celulari ai acestor neuroni sunt localizaţi în trunchiul cerebral, deşi fibrele care pornesc din ei (axonii) se termină în cortex. Acest lucru susţine principala funcţie de *activare* a acestor nuclei ai trunchiului cerebral, care împreună se numesc sistemul reticulat activator. Tocmai aceste căi activatoare erau lezate în cazul celor nouă pacienţi ai mei şi al sutelor de pacienţi supuşi leucotomiei documentaţi anterior care nu mai visau.

Începând din 1998, impulsionat şi de comentariile lui Braun despre implicaţiile descoperirii mele, mi-am îndreptat atenţia spre celelalte sisteme de activare din trunchiul cerebral.

Studiile cele mai interesante în acest domeniu le desfăşura Jaak Panksepp, a cărui lucrare enciclopedică *Afective Neuroscience* (1998) prezintă minuţios tot felul de dovezi în sprijinul opiniei sale că aceste sisteme presupuse a fi independente de conştiinţă, responsabile pentru reglarea doar a „nivelului” de conştiinţă,

generează un „conţinut” propriu.

Acest lucru s-a dovedit a fi foarte semnificativ.

*1 Sigmund Freud, *Studii despre isterie,* traducere din limba germană de Maria şi Ion Nastasia, cuvânt înainte de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2014, p. 242 (n.tr.).

Înainte de Freud şi după

În 1987 am mai luat o decizie care m-a pus în dezacord cu restul colegilor mei. M-am hotărât să mă specializez ca psihanalist[1]. Concluziile ce se conturau din studiile mele asupra viselor mă convinseseră că relatările subiective joacă un rol vital în neuropsihologie şi că opoziţia faţă de Freud dusese domeniul pe mai multe căi greşite. Dar nu rezultatele cercetărilor mele au fost factorul decisiv.

Ce m-a făcut să mă hotărăsc a fost un seminar la care am participat la University of the Witwatersrand, pe la mijlocul anilor 1980, coordonat de un profesor de literatură comparată pe nume Jean-Pierre de la Porte. Seminarul se referea la *Interpretarea viselor,* care îmi stârnise curiozitatea în contextul cercetării mele pentru doctorat. Ca toată lumea în acea perioadă, eram sceptic cu privire la Freud. Învăţasem din studenţie că psihanaliza este o „pseudoştiinţă”. Nimeni din domeniul ştiinţelor exacte nu îl mai lua în serios pe Freud, acesta fiind probabil motivul pentru care seminarul se desfăşura în cadrul unui departament de ştiinţe umaniste. Eu am participat pentru că Freud fusese

dispus să vorbească despre *conţinutul* viselor, tema mea de cercetare.

De la Porte a explicat că nu putem înţelege concluziile teoretice la care a ajuns Freud fără să-i analizăm mai întâi un manuscris anterior, scris în 1895, dar publicat abia în anii 1950, după moartea sa. Manuscrisul se intitulează „Proiect de psihologie”[2] şi în el Freud a încercat să-şi prezinte primele idei despre minte din perspectiva neuroştiinţelor.

Astfel, călca pe urmele marelui său profesor, fiziologul Ernst von Brücke, membru fondator al Societăţii de Fizică din Berlin. Misiunea acestei societăţi a fost formulată de Emil du Bois- Reymond în 1842 după cum urmează:

> Eu şi Brücke am jurat solemn să traducem în fapt acest adevăr: „În organism nu acţionează nicio altă forţă în afară de cele fizice şi chimice obişnuite. În cazurile ce nu pot fi explicate actualmente prin aceste forţe, trebuie fie să identificăm calea sau forma specifică a acţiunii lor prin metoda fizico-matematică, fie să presupunem că există forţe noi, egale ca importanţă cu forţele chimico-fizice inerente materiei, reductibile la forţa de atracţie şi de respingere”[3].

Iubitul lor profesor, Johannes Müller, se întrebase cum şi de ce viaţa organică diferă de materia anorganică. A concluzionat că „organismele vii sunt fundamental diferite de entităţile nonvii deoarece conţin un oarecare element nonfizic sau sunt guvernate de alte

principii decât lucrurile neînsufleţite”[4]. Pe scurt, conform lui Müller, organismele vii au o „energie vitală” sau „o forţă vitală” pe care legile fiziologice nu o pot explica. El considera că fiinţele vii nu pot fi reduse la mecanismele lor fiziologice componente deoarece sunt unităţi indivizibile cu *obiective* şi *scopuri,* pe care Müller le atribuia faptului că ele au un suflet. Dat fiind că termenul german *Seele* poate fi tradus fie prin „suflet”, fie prin „minte”[5], dezacordul dintre Müller şi studenţii lui seamănă izbitor de mult cu dezbaterea aprinsă din zilele noastre dintre filosofi ca Thomas Nagel şi Daniel Dennett cu privire la întrebarea dacă *conştiinţa* poate fi redusă la nişte legi fizice (Nagel pretinde că nu, Dennett pretinde că da).

Când am asistat la seminarul lui de la Porte, surprizător pentru mine a fost să aflu că Freud – deschizătorul de drumuri în ce priveşte investigarea subiectivităţii umane – se aliniase nu cu vitalismul lui Müller, ci cu fizicalismul lui Brücke. După cum a scris în primele rânduri ale „Proiectului” său din 1895: „[Este] intenţia de a oferi o psihologie ştiinţifică, adică de a prezenta procesele psihice ca pe nişte stări determinate cantitativ ale unor particule materiale detectabile”*[1] 6.

Nu realizasem că Freud gândea în termenii neuroştiinţelor. Acum aflam că abandonase cu reticenţă metodele neurologice de cercetare când a înţeles, cândva între 1895 şi 1900, că metodele disponibile pe atunci nu erau capabile să expună fundamentele fiziologice ale minţii.

Totuşi, schimbarea de atitudine a lui Freud a avut multe părţi bune. L-a forţat să analizeze mai atent

fenomenele psihologice ca atare şi să elucideze mecanismele funcţionale aflate la baza lor. Astfel a apărut modul psihologic de cercetare pe care l-a numit „psihanaliză”. Premisa fundamentală a acestuia era că fenomenele subiective manifeste (numite astăzi „explicite” sau „declarative”) au cauze latente (numite astăzi „implicite” sau „nondeclarative”). Adică Freud susţinea că fluxul haotic al gândurilor noastre conştiente poate fi explicat numai dacă presupunem că există conexiuni implicite între ele de care nu suntem conştienţi. De aici au apărut noţiunea de funcţii mintale latente şi celebra conjectură a lui Freud despre intenţionalitatea „inconştientă”.

Întrucât la sfârşitul secolului al XlX-lea nu existau metode de investigare a fiziologiei evenimentelor mintale inconştiente, mecanismele lor puteau fi deduse doar prin observaţie clinică. Ceea ce a aflat Freud astfel a generat a doua sa afirmaţie fundamentală. A observat că pacienţii adoptau o atitudine nici pe departe indiferentă faţă de intenţiile lor inconştiente deduse – mai degrabă, ei păreau că *nu vor* să le conştientizeze, şi nu că ar fi incapabili. Freud a folosit diverse denumiri pentru această tendinţă – „rezistenţă”, „cenzură”, „apărare” şi „reprimare” – şi a observat că previne detresa emoţională. Acest lucru, la rândul său, a relevat rolul central jucat de *sentimente* în viaţa psihică, modul în care susţin ele tot felul de biasuri cognitive. Aceste concluzii (evidente astăzi) i-au arătat lui Freud că unele dintre principalele forţe ce stimulează viaţa psihică sunt în întregime subiective, dar şi inconştiente. Investigarea lor sistematică l-a condus la a treia sa afirmaţie

fundamentală. Freud a concluzionat că în spatele sentimentelor stau în primul rând nevoile corpului, că, şi la oameni, şi la animale, viaţa psihică este *impulsionată* de imperativele biologice ale supravieţuirii şi reproducerii. Aceste imperative, considera Freud, constituie legătura dintre mintea care simte şi corpul fizic.

Freud a abordat într-un mod remarcabil de subtil relaţia minte-corp. A realizat că fenomenele psihologice pe care le studia nu erau direct reductibile la cele fiziologice. Încă din 1891 afirma că nu e posibil să atribui simptome psihologice unor procese neurofiziologice înainte de a reduce fenomenele psihologice şi fiziologice relevante (ambele părţi ale ecuaţiei) la *funcţiile* lor de bază. După cum am observat mai sus cu privire la prelucrarea informaţiilor, funcţiile pot fi executate la diverse niveluri[7]. Freud considera că psihologia şi fiziologia pot fi reconciliate numai pe terenul comun al funcţiei. Obiectivul său era să explice fenomenele psihologice prin intermediul unor legi funcţionale „metapsihologice” (termenul înseamnă „dincolo de psihologie”)[8]. Încercarea de a sări peste acest nivel funcţional al analizei, de a trece direct de la psihologie la fiziologie, se numeşte astăzi eroarea localizaţionistă[9].

În mod clar, pentru Freud, dacă nu şi pentru adepţii lui, psihanaliza trebuia să fie o etapă intermediară. Deşi, de la bun început, el a căutat să identifice legile care ne guvernează viaţa interioară bogată în experienţe subiective, viaţa psihică a rămas pentru el o problemă *biologică*[10]. După cum a scris în 1914, „toate

iterimatele noastre psihologice trebuie aşezate odată pe solul purtătorului organic”*[2] [11]. Prin urmare, a anticipat cu entuziasm ziua în care psihanaliştii se vor alătura din nou neuroştiinţelor:

> Într-adevăr, biologia constituie domeniul unor posibilităţi nelimitate, de la care aşteptăm explicaţiile cele mai surprinzătoare; deocamdată nu putem ghici răspunsurile pe care le va da, după câteva decenii [...]. Ele vor fi, poate, de natură să năruie întreaga noastră construcţie artificială de ipoteze*[3] [12].

Acesta nu era acel Freud extrem de speculativ despre care învăţasem în facultate. „Proiectul” era o revelaţie pentru mine, cum fusese şi pentru Freud. La vremea respectivă i-a scris prietenului său Wilhelm Fliess:

> Într-o noapte plină [...] barierele au fost brusc ridicate, vălurile au căzut şi am putut să înţeleg de la detaliile nevrozelor la determinanţii conştiinţei. Totul părea să se potrivească, rotiţele erau angrenate, obiectul îţi dădea impresia că era într-adevăr o maşină care, curând, va merge de la sine[13].

Dar euforia a durat puţin. O lună mai târziu scria: „Nu mai înţeleg în ce stare eram când am conceput «Psihologia», nu reuşesc să-mi dau seama cum am ajuns să v-o impun”[14]. Neavând metode neuroştiinţifice adecvate, Freud s-a bazat pe „imagini, transpoziţii şi

presupuneri” ca să-şi traducă inferenţele clinice mai întâi în termeni funcţionali, apoi în termeni fiziologici şi anatomici[15]. După o ultimă tentativă de revizuire (inclusă într-o lungă scrisoare trimisă lui Fliess pe 1 ianuarie 1896), „Proiectul” a dispărut din peisaj timp de vreo 50 de ani, după care a reapărut. Dar ideile pe care le conţine („fantoma sa ascunsă”, cum spunea James Strachey, cel care i-a tradus opera în engleză) i-au influenţat toate teoriile psihanalitice – în aşteptarea progresului ştiinţific viitor[16].

„Proiectul” conţinea două idei care ies în evidenţă în lumina descoperirilor contemporane. În primul rând, prozencefalul este un „ganglion simpatic” care monitorizează şi reglementează nevoile corpului. În al doilea rând, aceste nevoi sunt forţa motrice a vieţii psihice, „motorul mecanismului psihic”*[4] [17]. În lipsa oricărei perspective neurobiologice asupra modului în care asemenea nevoi corporale sunt reglate în creier – ca să nu mai vorbim de modul în care ar putea fi explicate „cu ajutorul metodei fizico-matematice” –, Freud nu a putut decât „să presupună că există forţe noi, egale ca
importanţă cu forţele chimico-fizice inerente materiei”, pentru a rămâne fidel idealurilor Societăţii de Fizică din Berlin. A numit „metapsihologice” aceste forţe *din spatele* fenomenelor psihologice. A precizat că dorea „să transforme metafizica în metapsihologie”*[5] [18]. Cu alte cuvinte, voia să înlocuiască filosofia cu ştiinţa – o ştiinţă a subiectivismului. Ne-a cerut să
nu îi judecăm prea aspru inferenţele speculative referitoare la procesele mintale latente:

> Această stare de lucruri se datorează numai faptului că am fost obligaţi să folosim anumiţi termeni ştiinţifici, adică limbajulul imagistic al psihologiei (mai exact, al psihologiei abisale). Fără ajutorul său nu am fi putut însă descrie procesele respective, mai mult, nici nu am fi putut să le identificăm. Probabil că minusurile descrierii noastre ar dispărea dacă am putea înlocui termenii psihologiei prin termeni din domeniul fiziologiei sau al chimiei*[6] 19.

Prima dintre noile forţe pe care Freud a fost obligat să le deducă era conceptul de „pulsiune”, pe care a definit-o drept „reprezentant psihic al stimulilor ce izvorăsc din interiorul corpului şi ajung în suflet, ca *o măsură a cererii de travaliu ce e impusă psihicului* ca urmare a legăturii sale cu ceea ce e corporal”*[7] 20.

Noţiunea lui Freud de „pulsiune” – pe care o considera sursa întregii „energii psihice” – nu era diferită de „energia vitală” a lui Müller, dar se înrădăcina în nevoile corpului. Freud a descris mecanismele cauzale prin care pulsiunile devin cogniţie intenţională drept o „economie a forţei nervoase”—. Totuşi, a recunoscut singur că „aici s-a petrecut ceva complet insesizabil pentru noi”*[8], cu referire la modul în care nevoile corpului devin energie psihică[22].

Când am citit aceste cuvinte, cam un secol mai târziu, am realizat că sosise vremea să înlocuim „termenii psihologiei cu termeni din domeniul fiziologiei sau al chimiei”. De pildă, forţa pulsională din spatele viselor, care era „latentă” în relatările subiective ale

pacienţilor lui Freud şi a cărei existenţă era, de aceea, considerată nefalsificabilă, era în mod clar „manifestă” în dovezile obiective obţinute prin metode fiziologice moderne in vivo care nu erau disponibile pe vremea lui Freud. Să luăm ca exemplu imaginile din figura 3, obţinute prin tomografie cu emisie de pozitroni[23]. Ele arată clar că circuitul CĂUTĂRII „plin de aspiraţii” se aprinde ca un pom de Crăciun în timpul somnului cu vise, în timp ce lobii prefrontali inhibitori nu sunt activi. În baza acestor descoperiri, când am fost invitat împreună cu Hobson la conferinţa Science of Consciousness din 2006 să discutăm despre credibilitatea ştiinţifică a teoriei visului elaborate de Freud, colegii noştri au votat cu 2:1 în favoarea ideii că această teorie este din nou viabilă[24].

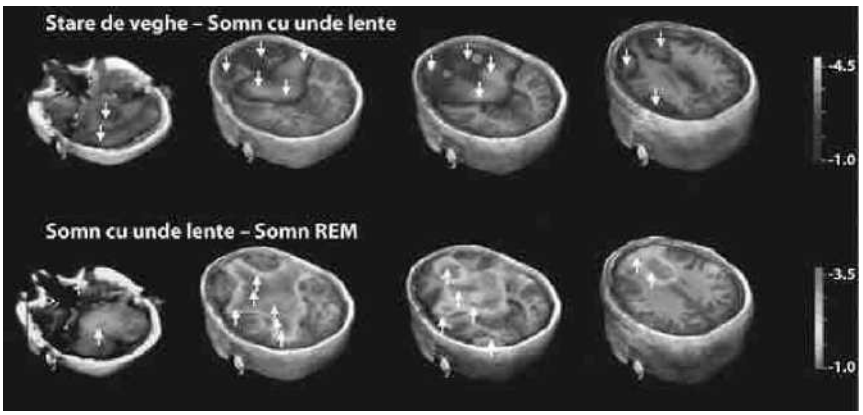


Figura 3. Rândurile orizontale arată secţiuni progresive în sus
ale creierului (de la stânga la dreapta). Rândul de sus arată
diferenţa dintre starea de veghe şi starea de somn, regiunea
umbrită descriind activarea corticală *diminuată* odată

cu
debutul somnului; rândul de jos arată diferenţa dintre somnul
REM şi somnul non-REM (cu unde lente), regiunea evidenţiată
descriind activarea subcorticală *crescută* odată cu debutul
somnului REM. Sistemul CĂUTĂRII se află în regiunea cu cea
mai intensă activare.

„Eul” subiectiv nu a fost niciodată exclus din psihanaliză, în pofida neajunsurilor sale. Avea cea mai importantă poziţie, oricât de stânjenitor era acest lucru pentru restul ştiinţei. Mulţi dintre colegii mei din mediul academic m-au sfătuit să nu asociez ceea ce făceam cu psihanaliza, având în vedere bagajul istoric al acestui cuvânt. Mi-au spus că era ca şi cum un astronom s-ar fi asociat cu astrologia. Dar eu am considerat că nu ar fi onest din punct de vedere intelectual dacă nu i-aş recunoaşte meritele lui Freud. Indiferent cât de parţial şi-a atins obiectivele, ele erau în mod clar *obiectivele* corecte pentru o ştiinţă a minţii. Aşa că mi-am denumit abordarea „neuropsihanaliză”. Am spus că neuropsihologia care mi s-a predat la cursuri putea foarte bine să se fi numit neurobehaviorism, datorită atitudinii sale faţă de subiectivitate. Am vrut să fie clar că neuropsihologia pe care o elaboram se baza pe experienţa trăită. În acest spirit, după ce am scris o lucrare programatică despre relaţia dintre psihanaliză şi neuroştiinţe, m-am pus pe

treabă[25].

M-am mutat în Londra în 1989, ca să fac studii de psihanaliză. Ca să-mi continui cercetările şi activitatea clinică, am acceptat şi poziţia de *honorary lecturer* în neurochirurgie în cadrul Royal London Hospital School of Medicine. Am fost încântat să-i urmez puternica tradiţie în neurologie: la mijlocul secolului al XIX-lea, John Hughlings Jackson, părintele fondator al neurologiei şi neuropsihologiei britanice, profesa acolo ca medic. Spitalul Royal London se afla în Whitechapel, o zonă ce îi atrage de secole pe imigranţi, deservind, astfel, dintotdeauna comunităţi vulnerabile. Îmi amintea de spitalul Baragwanath din Soweto. Mi-a fost ca o a doua casă.

La începutul anilor 1990, un coleg neurochirurg din Africa de Sud mi-a prezentat cazul domnului S., un pacient pe care, cu zece luni înainte, îl operase pentru a-i extirpa o tumoră ce îi creştea sub lobii frontali, deplasându-i nervii optici. Domnul S. suferise o mică hemoragie în timpul intervenţiei, care îi întrerupsese fluxul sangvin spre prozencefalul bazal (vezi figura 1). Nucleii prozencefalului bazal transmit acetilcolina către diversele structuri corticale şi subcorticale implicate în accesarea memoriei pe termen lung. Se consideră că aceste căi colinergice interacţionează cu căile dopaminergice (vezi figura 2), acestea din urmă fiind aşa-numitul sistem de recompensă ce activează comportamentele de „căutare”, nu numai în legătură cu acţiunile fizice din lumea exterioară, ci şi cu lumea interioară a reprezentărilor, acţiunile imaginare ce apar în gândire şi în vise[26]. Din cauza acestei hemoragii,

domnul S. s-a trezit din operaţie cu un sindrom amnezic profund, numit psihoză Korsakoff, a cărei caracteristică principală este o stare asemănătoare visului numită confabulaţie. Memoria sa legată de evenimentele recente era profund dereglată, astfel încât avea constant amintiri false. Acest deficit de *căutare* este destul de debilitant, dar în cazul amneziei confabulatorii este agravat de faptul că pacienţii nu îşi *monitorizează* adecvat veridicitatea amintirilor pe care le accesează în mod eronat, astfel tratându- le ca şi cum ar fi reale, când, evident, nu sunt.

De exemplu, domnul S. credea că se află în Johannesburg (unde locuia), dar, de fapt, tocmai călătorise la Londra, ca să îl consult. Nu îşi amintea să fi călătorit. Când l-am corectat în această privinţă, a insistat că nu avea cum să fie în Londra. Aşa că l-am rugat să se uite pe fereastră, întrucât ningea, iar asta nu se întâmplă niciodată în Johannesburg. Iniţial a părut şocat, apoi s-a liniştit şi a replicat: „Nu, *ştiu* că sunt în Johannesburg – dacă mănânci pizza nu înseamnă că şi eşti în Italia”.

Domnul S. era inginer electronist şi avea 56 de ani. L-am consultat zilnic la clinica mea pentru tratament ambulatoriu de zi, de şase ori pe săptămână, în încercarea de a-l orienta şi a-l ajuta să înţeleagă modul în care îl trăda memoria. Deşi îl consultam zilnic în acelaşi loc şi la aceeaşi oră, nu m-a recunoscut niciodată drept medicul lui de la o întâlnire la alta. Aparent, îmi recunoştea chipul, dar mă confunda de multe ori cu persoane pe care le ştia din alte contexte – de obicei cu un coleg inginer care lucra cu el la vreo

problemă de electronică sau cu un client care avea nevoie de ajutorul lui de specialitate. Cu alte cuvinte, domnul S. mă trata ca şi cum eu aş fi avut nevoie de ajutorul lui, nu invers. De asemenea, îşi imagina frecvent că eram amândoi studenţi şi că ieşiserăm la un pahar după vreo activitate sportivă (un concurs de canotaj sau un meci de rugbi). La vremea respectivă eu eram suficient de tânăr ca aşa ceva să fie plauzibil, dar domnul S. terminase facultatea de mai bine de 30 de ani.

După fiecare consult clinic, discutam separat cu soţia lui ca să contextualizez amintirile lui eronate şi să încerc să le găsesc semnificaţia. Aceasta era principala diferenţă dintre abordarea mea şi cea a colegilor mei, mai tradiţională, faţă de „reabilitarea cognitivă”. Neuropsihologii sunt preocupaţi de obicei de *gradul* tulburării de memorie, evaluat din perspectiva persoanei a treia, însă pe mine mă interesa mai mult *conţinutul* subiectiv al erorilor domnului S., înţeles din perspectiva persoanei întâi. Am plecat de la ipoteza că semnificaţia personală a evenimentelor ce îi veneau obsesiv în minte, în locul amintirilor-ţintă pe care le căuta, avea să clarifice cât de cât mecanismul acestor confabulaţii, deschizând astfel căi noi spre influenţarea lor. Aşadar, la întâlnirile mele cu soţia lui, de exemplu, voiam să ştiu dacă domnul S. chiar făcuse parte din vreo echipă de canotaj şi de rugbi în studenţie şi dacă oferea într-adevăr asistenţă de specialitate în probleme de electronică.

Două lucruri pe care le-am aflat astfel sunt relevante pentru înţelegerea confabulaţiilor sale. Mai întâi, e

vorba de faptul că suferise cândva de nişte probleme cronice la dantură – probleme tratate (cu succes) în cele din urmă, cu ajutorul unor implanturi dentare – şi, în al doilea rând, că suferea de aritmie cardiacă, ţinută sub control cu ajutorul unui stimulator cardiac.

Am selectat un scurt fragment din transcrierea unei înregistrări audio a primelor câteva minute din a zecea consultaţie pe care am avut-o cu domnul S. Am ales acest fragment deoarece, când m-am dus să-l preiau din sala de aşteptare în ziua aceea, a părut pentru scurt timp (şi pentru prima dată) că recunoştea cine sunt şi de ce venise la mine la consultaţie. Când am intrat în sala de aşteptare, şi-a atins cicatricea de la craniotomie din vârful capului şi a spus: „Bună, doctore”.

Ne-am instalat în cabinetul meu, eu sperând să aflu mai multe, profitând de acest licăr de introspecţie, dacă despre aşa ceva era vorba.

Eu: Aţi dus mâna la cap când ne-am văzut în sala de aşteptare.
DomnulS.: Cred că problema e că lipseşte un cartuş. Trebuie să... ne trebuie doar specificaţiile. Ce era? Un C49? Ar trebui să-l comandăm?
Eu: Ce face un cartuş C49?
DomnulS.: Memorie. E un cartuş de memorie, un implant de memorie. Dar n-am înţeles niciodată cum funcţionează. De fapt, nu l-am mai folosit de vreo cinci-şase luni. Se pare că n- avem nevoie de el de fapt. A fost scos tot de un doctor. Cum îl cheamă? Cred că Solms. Dar se pare că nu-mi trebuie de fapt. Implanturile funcţionează bine.

Eu: Vă daţi seama că ceva nu e în regulă cu memoria dumneavoastră, dar.

Domnul S.: Da, nu funcţionează 100%, dar nu ne trebuie de fapt – doar avea o mică defecţiune, îi lipseau nişte părţi. Analiza arată că lipsea un C sau un C09. Denise [prima lui soţie] m-a adus aici să mă vadă un doctor. Cum îl cheamă? Iar am uitat. Solms sau aşa ceva. Şi a făcut un transplant din ăla de inimă şi acum merge bine, nu ratează nicio bătaie.

Eu: Vă daţi seama că ceva lipseşte. Unele amintiri lipsesc şi, desigur, asta vă îngrijorează. Speraţi că eu pot să rezolv problema, aşa cum ceilalţi doctori v-au rezolvat problemele cu dantura şi cu inima. Dar vă doriţi acest lucru atât de mult, încât vă e greu să acceptaţi faptul că încă nu s-a rezolvat.

Domnul S.: A, înţeleg. Da, nu merge 100%. [Îşi atinge capul.]

M-am lovit la *c*ap. Am ieşit de pe teren câteva minute. Dar acum e bine. Presupun că n-ar trebui să mă întorc. Dar mă ştii doar, nu-mi place să stau pe tuşă. Aşa că l-am întrebat pe Tim Noakes [celebru medic sportiv din Africa de Sud] – că asigurarea medicală o am, ştii, aşa că de ce să n-o folosesc, de ce să nu mă duc la cel mai bun – şi el a spus: „Bine, joacă mai departe”.

Mă opresc aici. Ar trebui să fie suficient de uşor de recunoscut tulburările pur cognitive ale căutării şi monitorizării memoriei pe care le-am menţionat mai sus. Când domnul S. m-a văzut intrând în sala de aşteptare pentru cea dea zecea consultaţie, apariţia

mea i-a evocat o mulţime de asocieri – legate de doctori, capul său, amintirile lipsă, intervenţiile chirurgicale şi altele asemenea. Dar în niciunul dintre aceste cazuri nu a accesat exact amintirea-ţintă pe care o căuta. În schimb, au apărut cele ce pot fi caracterizate drept „ţinte potenţiale” – amintiri din aceleaşi categorii semantice largi ca şi ţintele, dar localizate greşit în timp şi spaţiu. Astfel, ideea de „doctor” a evocat asocieri referitoare la un neurochirurg şi la un celebru medic sportiv în locul ţintei, eu; ideea de „cap” a evocat o situaţie când s-a lovit la cap în locul unei tumori cerebrale; „amintirea lipsă” a evocat un cartuş electronic în locul amneziei lui; „intervenţiile chirurgicale” i-au evocat intervenţiile dentare şi cardiologice anterioare în locul intervenţiei chirurgicale recente pe creier şi aşa mai departe. Deficitul de monitorizare e la fel de simplu de observat: domnul S. a acceptat veridicitatea amintirilor sale eronate cu prea multă uşurinţă. Faptul că se simţea ca un student de douăzeci şi ceva de ani pe un teren de rugbi (în pofida tuturor dovezilor contrare) este un exemplu evident în acest sens. La fel şi convingerea lui că încă se afla în Johannesburg.

Dar, când analizăm confabulaţiile domnului S. dintr-un punct de vedere *subiectiv,* apar şi alte fapte. Imaginează-ţi cum te-ai *simţi* dacă ai realiza brusc că nu îl recunoşti pe medicul care tocmai a intrat în încăpere, deşi el pare să fie responsabilul pentru serviciile medicale pe care le primeşti, că nu ştii în ce încăpere (sau chiar în ce oraş) te afli, că ai o cicatrice mare în vârful capului şi nu ştii cum a apărut acolo, că – de fapt

– nu îţi aminteşti ce s-a întâmplat cu doar două minute în urmă, ca să nu mai vorbim de zilele şi lunile dinaintea momentului prezent. Probabil ai simţi un soi de *panică,* întrebându-te dacă nu cumva acest medic ţi-a făcut o operaţie la cap, în urma căreia nu îţi mai aminteşti nimic de la un moment la altul. Iată cum resimte subiectul intenţional al minţii – eul viu – lipsa mecanismelor de căutare şi de monitorizare a amintirilor.

Să observăm acum ce a făcut domnul S. ca urmare a acestor sentimente (cu alte cuvinte, să observăm efectele lor de cauzalitate asupra cogniţiei lui). Când a realizat că îi lipseşte „cartuşul de memorie”, s-a liniştit (printr-un delir) că *poţi pur şi simplu să-ţi comanzi unul nou.* Pentru că nu a fost cu totul convins de propriile argumente, s-a răzgândit. De fapt, *nu ai neapărat nevoie de cartuş,* poţi să te descurci şi fără, iar el deja se descurcă fără probleme de luni de zile. Apoi a făcut o legătură între cartuşul lipsă şi cicatricea de la craniotomie: se pare că un doctor i-a tăiat ceva. Speră că nu e vorba de doctorul care stă în faţa lui şi, mai mult, speră că intervenţia nu a fost un eşec. În acest moment, domnul S. îşi aminteşte că intervenţia chirurgicală dentară şi cea cardiacă prin care a trecut au avut succes şi contopeşte (printr-un delir) aceste intervenţii cu cea de la cap: *a fost reuşită, implanturile funcţionează bine* şi „nu ratează nicio bătaie”. Când introduc o sămânţă de îndoială în partitura lui, schimbă direcţia. E de acord că nu funcţionează 100%, dar, în acelaşi timp, hotărăşte că ceea ce i s-a întâmplat la cap nu a fost intervenţie chirurgicală până la urmă, *ci doar*

s-a lovit – acum este afectat de efectele temporare ale unei accidentări minore pe terenul de sport. În consecinţă, a fost trimis pe tuşă câteva minute. Dar, din fericire, pentru că îşi permite să achite o consultaţie la cel mai bun medic sportiv, se linişteşte din nou: *poate juca mai departe.* Totul va fi bine.

Analizarea confabulaţiilor domnului S. din perspectiva persoanei întâi ne relevă în mod clar ceva nou despre ele: conţinutul amintirilor sale eronate are un caracter puternic *motivat.* Ele sunt departe de a fi nişte erori întâmplătoare de căutare, conţin un bias cognitiv clar – au obiectivul şi scopul de a-i reîncadra situaţia încărcată de anxietate drept una liniştitoare, sigură şi familiară. Aşadar, după cum a dedus şi Freud în cazul viselor, confabulaţiile sunt motivate. Procesele psihice din amnezia confabulatorie vin din nişte *aspiraţii*. Dar acest fapt devine evident numai atunci când contextul emoţional şi semnificaţia personală (trăită numai de domnul S.) ale implanturilor dentare („implanturile funcţionează bine”) şi stimulatoarelor cardiace („nu ratează nicio bătaie”) sunt luate în considerare – cum ar proceda un psihanalist. Iată ce nu reuşesc să vadă neuropsihologii când urmăresc să fie cu totul obiectivi; cum spunea Sacks, când exclud psihicul.

Perspectiva observaţiei la persoana întâi pe care tocmai am descris-o relevă şi ceva nou despre *mecanismul* confabulaţiei, ceva trecut cu vederea din perspectiva persoanei a treia. Ea ne spune că confabulaţia nu apare exclusiv din cauza deficitelor de căutare strategică şi de monitorizare a surselor („cartuşele de memorie” lipsă), ci şi din cauza

dezinhibării unor forme de reamintire mediate într-un mod mai emoţional, cam aşa cum funcţionează memoria unui copil. Acest mecanism psihodinamic are implicaţii pentru tratamentul confabulaţiei şi, desigur, pentru întrebarea referitoare la procesele cerebrale pe care le implică ea. Se dovedeşte că funcţiile corecte de căutare şi de monitorizare ale memoriei depind parţial de circuitele colinergice din prozencefalul bazal, care constrâng mecanismele de „recompensă” ale circuitului dopaminergic mezocortical-mezolimbic în accesarea memoriei. Se întâmplă că în cazul viselor apare o eliberare similară a căutării dopaminergice—. Din acest motiv, le-am prezentat colegilor mei cazul domnului S. sub titlul „Omul care trăia într-un vis”.

Acest lucru mi-a permis, ca în cazul viselor, să leg provizoriu mecanismul dopaminergic neîngrădit al „recompensei”, al „dorinţei” sau al „CĂUTĂRII” de noţiunea lui Freud de „împlinire a dorinţelor”[28] – un concept metapsihologic strâns legat de conceptul său de „pulsiune”[29]. Dimpotrivă, funcţiile nucleilor colinergici din prozencefal pot fi legate din anumite puncte de vedere de influenţele inhibitorii ale „testării realităţii”[30]. În acest fel, am început să traduc deducţiile lui Freud despre mecanismele funcţionale ale subiectivităţii prin echivalentele lor fiziologice.

Aceştia au fost primii mei paşi. În mod firesc, asemenea generalizări ample nu se pot baza pe dovezi pur clinice dintr-un singur caz. Prin urmare, după ce mi-am formulat impresia despre domnul S., am cooptat evaluatori „în orb” (colegi nefamiliarizaţi cu ipoteza mea) pentru a măsura, pe o scală Likert în şapte

puncte, gradul de agreabilitate versus dezagreabilitate într-un eşantion continuu neselectat alcătuit din 155 dintre confabulaţiile pacientului. Rezultatele au fost (foarte) semnificative statistic: când erau comparate cu amintirile-ţintă pe care le-au înlocuit, confabulaţiile domnului S. îi îmbunătăţeau substanţial situaţia din punct de vedere emoţional[31]. În continuare, am demonstrat împreună cu colaboratorii mei acelaşi efect puternic în studiile în care erau implicaţi numeroşi alţi pacienţi cu confabulaţii. Studiile empirice ulterioare au validat statistic efectele confabulaţiei de reglare a dispoziţiei afective pe care le-am dedus clinic în cazul domnului S.[32]. Acest program de cercetare a iniţiat o abordare cu totul nouă a neuropsihologiei confabulaţiei[33] şi a tulburărilor conexe precum anosognozia[34]. De asemenea, a pus bazele unei abordări noi a tulburărilor psihiatrice comune precum dependenţa şi depresia majoră[35]. De trei decenii lucrez la dezvoltarea acestei abordări „neuropsihanalitice” a bolilor mintale, încercând să readuc subiectivitatea în neuroştiinţe[36].

Pe măsură ce am acumulat experienţă clinică de tipul celei tocmai descrise în timpul specializării mele în psihanaliză la Londra, am fost invitat să-mi expun constatările într-o serie de prezentări ştiinţifice la New York. Acestea au început în 1992 cu un simpozion de o zi la Academia de Medicină din New York şi au continuat sub forma unor seminare lunare cu cei mai apropiaţi colaboratori ai mei la Societatea şi Institutul de Psihanaliză din New York[37]. Treptat, întâlnirile

acestui grup de colegi au generat activităţi similare şi în alte părţi ale lumii, ducând la decizia (în 1999) de a înfiinţa o revistă nouă care să servească drept vehicul pentru comunicarea dintre noi. Revista avea nevoie de un nume şi aşa am folosit pentru prima dată termenul inventat de mine, *Neuropsychoanalysis.*

Activitatea mea în acest domeniu interdisciplinar a primit încurajări puternice din partea lui Eric Kandel, pe care l-am cunoscut în 1993. Kandel este atipic printre specialiştii în neuroştiinţe prin marele său respect pentru Freud. De fapt, iniţial intenţionase să se specializeze în psihanaliză, dar a fost descurajat de Ernst Kris, un analist important la momentul respectiv şi tatăl iubitei de atunci a lui Kandel. Acesta din urmă nu ascunde faptul că Kris considera că nu i se potrivea practica clinică în psihiatrie şi cred că îi este recunoscător că l-a îndrumat spre cercetare în domeniu.

La cinci ani după ce l-am cunoscut (şi cu doi ani înainte să i se acorde Premiul Nobel), Kandel a publicat un articol intitulat „A new intellectual framework for psychiatry”, în care afirma că psihiatria secolului XXI ar trebui să se bazeze pe integrarea neuroştiinţelor şi psihanalizei[38]. Într-un articol ulterior, a afirmat: „Psihanaliza încă reprezintă viziunea asupra minţii cea mai coerentă şi mai satisfăcătoare din punct de vedere intelectual”[39]. Această opinie coincidea cu a mea: în pofida tuturor neajunsurilor sale, psihanaliza constituie în prezent cel mai bun punct de plecare sub aspect conceptual pentru abordarea ştiinţifică a subiectivităţii.

Aşadar, nu m-a surprins când Kandel mi-a acceptat invitaţia de a face parte din consiliul editorial fondator

al revistei *Neuropsychoanalysis,* alături de o masă critică de alţi specialişti în neuroştiinţe şi psihanalişti importanţi, care au fost de acord că acesta era drumul corect pe care trebuiau să-l urmeze disciplinele noastre[40]. Un an mai târziu, la începutul noului secol, am fondat Societatea Internaţională de Neuropsihanaliză, cu Jaak Panksepp şi cu mine ca primi copreşedinţi. Acest eveniment a coincis cu congresul inaugural al societăţii, care, de atunci, se ţine anual în diferite oraşe din lume. Tema primului congres au constituit-o emoţiile. Întâlnirea a avut loc la Royal College of Surgeons of England, invitaţii fiind Oliver Sacks, Jaak Panksepp, Antonio Damasio şi eu.

Am menţionat deja relaţia mea cu Oliver Sacks. Am vorbit şi despre lucrarea lui Jaak Panksepp *Afective Neuroscience,* al cărei titlu face aluzie la faptul că neuroştiinţa cognitivă nu acorda atenţie suficientă „afectului” (termenul tehnic pentru sentimente). După ce am citit-o, am stabilit o colaborare ştiinţifică strânsă cu Panksepp, ceea ce a deplasat decisiv centrul atenţiei în activitatea mea în următoarele două decenii de la cortex la trunchiul cerebral. Îi sunt profund îndatorat pentru că mi-a arătat calea spre ideile despre care voi vorbi în paginile următoare, acesta fiind motivul pentru care îi dedic cartea de faţă.

M-am familiarizat mai întâi cu lucrările lui Antonio Damasio şi ale partenerei sale, Hanna Damasio, în timpul studiilor mele de neuropsihologie. Erau nişte specialişti în neuroştiinţele cognitive foarte respectaţi, al căror manual, *Lesion Analysis in Neuropsychology* (1989), a fost indispensabil pentru cercetările mele

asupra viselor. Totuşi, ceea ce l-a făcut pe Damasio cunoscut în toată lumea a fost volumul său *Eroarea lui Descartes* (1994), o pledoarie înflăcărată adresată neuroştiinţei cognitive pentru o mai mare valorizare a afectului.

De asemenea, Panksepp şi Damasio au jucat un rol important în cadrul celui de-al Doisprezecelea Congres Internaţional de Neuropsihanaliză din 2011, care s-a dovedit a fi un punct de cotitură în domeniu. Congresul s-a desfăşurat la Berlin, pe tema creierului in vivo. Ceilalţi invitaţi au fost Bud Craig şi Vittorio Gallese – doi dintre principalii experţi din lume pe acest subiect. De obicei, rolul meu la aceste congrese este să prezint observaţiile finale, ce rezumă temele principale şi, mai important, *integrează* perspectivele din neuroştiinţe şi psihanaliză prezentate. De data aceasta, sarcina mea a fost neobişnuit de dificilă.

Prima dificultate venea din faptul că Damasio se contrazisese puternic cu Craig în timpul congresului cu privire la modul în care apare în creier „sinele” care simte. Deşi amândoi erau de acord că simţul sinelui apare din regiunile cerebrale care monitorizează starea corpului visceral, Damasio – în spiritul lui Panksepp – considera că mecanismele în cauză se găsesc cel puţin parţial în trunchiul cerebral. Craig, dimpotrivă, pretindea că se localizează exclusiv în cortex – în insula anterioară, mai exact. Acest dezacord a fost relativ simplu de rezolvat în concluziile mele de final, deoarece Damasio furnizase date convingătoare, axându-se pe un pacient căruia i se îndepărtase complet cortexul insular. Voi descrie acest pacient în capitolul următor.

Mult mai dificil de reconciliat a fost o contradicţie majoră apărută în timpul congresului între perspectivele noi ale lui Panksepp şi Damasio, pe de o parte, şi cele vechi ale lui Freud, pe de altă parte.

Pacientul lui Damasio fără cortexul insular „raporta senzaţii de foame, sete şi dorinţa de a excreta şi se comporta în consecinţă”[41]. Aceste senzaţii sunt exemple a ceea ce Panksepp numeşte „afecte homeostatice” – afecte ce reglează nevoile vitale ale corpului. Freud le numea „pulsiuni” – sursa „energiei psihice”, „motorul mecanismului psihic”. Termenul general al lui Freud pentru partea minţii care îndeplineşte aceste funcţii vitale era „Se”:

> Se-ul, separat de lumea exterioară, îşi are propria sa lume perceptivă. El simte cu o acuitate extraordinară anumite transformări din interiorul său, mai cu seamă oscilaţiile din tensiunea nevoilor a pulsiunilor lui, care devin conştiente ca senzaţii ale seriei plăcere-neplăcere. Este, ce-i drept, greu de indicat pe ce căi şi cu ajutorul căror organe terminale sensibile iau naştere aceste percepţii. Dar este cert că autopercepţiile – afecte generale şi senzaţii de plăcere- neplăcere – domină procesele din Se cu o forţă despotică. Seul ascultă de principiul neînduplecat al plăcerii*[9] 42.

După cum am spus, obiectivul meu ştiinţific era să traduc asemenea noţiuni metapsihologice în limbajul anatomiei şi al fiziologiei, astfel încât să pot integra în

neuroştiinţe abordarea lui Freud. Dar aici m-am poticnit într-o contradicţie majoră a concepţiei clasice a lui Freud: el ajunsese la concluzia că „Se-ul” este inconştient. Aceasta a fost una dintre concepţiile lui fundamentale cu privire la modul în care funcţionează mintea. Îmi era clar că partea creierului ce măsoară „cererea de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce e corporal” – partea care generează ceea ce Freud numea „pulsiuni”, care sunt sinonime cu „afectele homeostatice” ale lui Panksepp (ce declanşează mecanismul lui de CĂUTARE venind din aspiraţii) – se afla în trunchiul cerebral şi în hipotalamus (vezi figura 1). Aceasta este partea creierului care se supune „principiului plăcerii”. Dar cum pot fi inconştiente senzaţiile de plăcere? După cum am văzut la

pacientul lui Damasio, pulsiunile precum foamea, setea şi dorinţa de a excreta sunt *simţite.* Desigur că aşa e. Totuşi, Freud a spus că Se-ul – sediul pulsiunilor – este inconştient. Asimilase aceeaşi doctrină clasică precum Craig (ca şi mine, cel puţin iniţial) şi, prin urmare, localizase conştiinţa în cortexul cerebral. Aşa că a scris următoarele în eseul său din 1920 pe care l-am citat mai sus şi în care spera că neajunsurile din teoriile sale vor dispărea odată ce vom putea înlocui termenii psihologici cu termeni din fiziologie şi din chimie:

> Cum conştiinţa realizează în special percepţii ale unor excitaţii care provin din lumea exterioară şi senzaţii de plăcere sau neplăcere, care nu-şi pot avea

sursa decât în interiorul aparatului psihic, este posibilă o situare spaţială a sistemului *P-Cs* (percepţie-conştiinţă). El trebuie să se afle la graniţa care separă lumea exterioară de cea interioară, să fie orientat către lumea exterioară şi să înglobeze toate celelalte sisteme psihice. Observăm însă că, făcând aceste presupuneri, nu am spus nimic nou, ci *ne-am raliat anatomiei localizante a creierului care situează „sediul" conştiinţei în scoarţa cerebrală,* stratul cel mai exterior, periferic al organului central. Anatomia creierului nu are motive să se întrebe de ce, anatomic vorbind, conştiinţa este localizată tocmai la suprafaţa creierului, şi nu într-o regiune mai protejată, undeva în straturile intime ale acestuia*[10] 43.

În caz că mai există vreo îndoială în legătură cu opinia lui Freud că *toată* conştiinţa (inclusiv senzaţiile de plăcere şi neplăcere) este localizată în cortex, mai dau un citat:

Conştientizarea este legată mai cu seamă de percepţiile pe care organele noastre senzoriale le capătă de la lumea exterioară. Este deci pentru consideraţia topică un *fenomen care se desfăşoară în stratul cel mai exterior al scoarţei Eului.* Noi primim, desigur, şi ştiri conştiente din interiorul corpului, sentimentele, care chiar influenţează substanţial viaţa noastră psihică decât percepţiile exterioare şi, în anumite împrejurări, transmit şi organelor senzoriale sentimente, senzaţii de durere, în afară de percepţiile lor specifice. Dar, întrucât aceste senzaţii, cum se

numesc ele spre deosebire de percepţiile conştiente, emană tot de la organele terminale şi *noi le concepem pe toate ca prelungiri, ca derivate ale stratului cortical,* putem menţine în picioare afirmaţia de mai sus. Deosebirea ar fi numai că, în ceea ce priveşte organele terminale ale senzaţiilor şi sentimentelor, însuşi corpul ar înlocui lumea exterioară[*11] 44.

În mod clar, pentru Freud, sentimentele conştiente, nu mai puţin decât percepţiile, sunt generate în „Eu” (acea parte a minţii pe care a identificat-o cu cortexul)[45], şi nu în „Se-ul” inconştient – pe care eram acum obligat să-l localizez în trunchiul cerebral şi hipotalamus. Pe scurt, părea că Freud a înţeles invers relaţia funcţională dintre „Se” (trunchiul cerebral) şi „Eu” (cortex), cel puţin în ce priveşte sentimentele. El credea că Eul care percepe este conştient, iar Se-ul care simte este inconştient. Se poate oare să fi înţeles el pe dos modelul minţii[46]?

31 (n.tr.).

*2 Sigmund Freud, „Pentru a introduce narcisismul”, traducere de Gilbert Lepădatu, în *Opere esenţiale,* vol. 3: *Psihologia inconştientului,* notă asupra ediţiei de Raluca Hurduc, note introductive de Roxana Melnicu. Bucureşti: Trei, 2017, ebook (n.tr.).

*3 Sigmund Freud, „Dincolo de principiul plăcerii”, în

*1 Sigmund Freud, *Opere,* vol. XVII: *Proiect de psihologie,* traducere din limba germană de Daniela Ştefănescu, cuvânt înainte de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2007, p.

Opere esenţiale, vol. 3: *Psihologia inconştientului,* traducere de G. Purdea şi Vasile Dem. Zamfirescu. Bucureşti: Trei, e-book (n.tr.).

*4 Sigmund Freud, *Proiect de psihologie,* ed.cit., pp. 41, 55 (n.tr.).

*5 Sigmund Freud, *Psihopatologia vieţii cotidiene. Despre uitare, greşeli de vorbire, superstiţie şi eroare,* traducere de Herta Spuhn şi Daniela Ştefănescu, ediţia a Il-a. Bucureşti: Trei, 2010, e-book (n.tr.).

*6 Sigmund Freud, „Dincolo de principiul plăcerii", ed.cit. (n.tr.). *7 Sigmund Freud, „Pulsiuni şi destine ale pulsiunilor", în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, ed.cit. (n.tr.).

*8 Sigmund Freud, „Compendiu de psihanaliză", în *Despre psihanaliză,* traducere din germană de Daniela Ştefănescu, prefaţă de Vasile Dem Zamfirescu. Bucureşti: Trei, 2014, p. 241 (n.tr.).

*9 *Ibidem,* p. 243 (n.tr.).

*10 Sigmund Freud, „Dincolo de principiul plăcerii", ed.cit. (n.tr.).

*11 *Ibidem,* p. 193 (n.tr.).

Eroarea privitoare la cortex

Pe la sfârşitul anului 2004, Björn Merker, un specialist în neuroştiinţe, a plecat împreună cu cinci familii cu copii cu afecţiuni neurologice într-o excursie de o săptămână la Disney World. Copiii aveau vârste între zece luni şi cinci ani. Au ales diverse atracţii: una dintre preferate a fost „It's a Small World After All". Au făcut poze cu Mickey Mouse. Au mâncat popcorn, crenvurşti şi îngheţată – poate mai mult decât era cazul – şi au băut mult suc. Uneori, păreau că sunt copleşiţi, au plâns de multe ori. Dar, în pofida acestor momente stresante, doctorul Merker a fost frapat de cât de mult păreau să se distreze copiii – cât de mult le plăcea că sunt acolo. La urma urmei, conform uneia dintre premisele fundamentale ale neurologiei, aceşti copii ar fi trebuit să fie într-o „stare vegetativă". Ar fi trebuit să fie la un pas de comă: capabili să manifeste doar funcţii autonome precum reglarea bătăilor inimii, respiraţia şi activitatea gastrointestinală, răspunsurile lor motoare limitându-se la simple reflexe precum clipitul sau înghiţitul.

Merker se împrietenise cu aceste familii în anul

precedent, când s-a alăturat unui grup de întrajutorare pentru îngrijitorii copiilor care suferă de o afecţiune cerebrală rară numită hidroanencefalie. De pe la începutul lui 2003, Merker citise peste 26.000 de e-mailuri schimbate între membrii grupului. Având argumente teoretice, se îngrijorase că exista posibilitatea ca presupunerea generală că aceşti copii vor fi „vegetativi” să fi devenit o profeţie care se autoîmplineşte nu din cauza afecţiunii în sine, ci a faptului că asemenea pacienţi erau *trataţi* de majoritatea pediatrilor şi neurologilor ca şi cum ar fi fost complet insensibili. Iată ce a raportat:

> Aceşti copii sunt nu numai treji şi adesea vioi, ci manifestă şi receptivitate la mediul ambiant sub forma unor reacţii emoţionale sau de orientare la evenimente din mediu [...]. Mai mult, copiii au epilepsie cu crize de absenţă. Părinţii recunosc aceste întreruperi de accesibilitate la copiii lor, comentând despre ele în termeni precum „ia o pauză să vorbească cu îngerii” şi dându-şi seama foarte uşor când copilul „şi-a revenit”. [...] Faptul că aceşti copii prezintă asemenea episoade pare să fie o dovadă serioasă a stării lor de conştiinţă[1].

Cea mai importantă observaţie a lui Merker nu este că pacienţii îşi pierd şi îşi recapătă luciditatea, ci că manifestă „receptivitate la mediul ambiant sub forma unor reacţii emoţionale sau de orientare la evenimente din mediu”. Aceasta este trăsătura definitorie a ceea ce se presupune că le lipseşte pacienţilor vegetativi:

intenţionalitatea. Din acest motiv, starea vegetativă este definită şi ca „veghe *fără reacţie”*[2]*.* În general, se consideră că aceşti pacienţi înregistrează stimuli vizuali, auditivi şi tactili *în mod inconştient,* dar Merker i-a văzut pe copii exprimându-şi plăcerea prin zâmbete şi râsete, iar aversiunea prin proteste, arcuiri ale spatelui şi plânsete, „chipurile fiindu-le animate de aceste stări emoţionale”. I-a observat pe adulţii familiari copiilor folosindu-se de reacţiile lor pentru a construi secvenţe de joacă, copiii progresând previzibil de la zâmbete la chicoteli, râsete şi mult entuziasm. Copiii răspundeau cel mai puternic la vocile şi acţiunile părinţilor şi ale altor persoane cu care erau familiarizaţi şi manifestau preferinţe pentru anumite situaţii în detrimentul altora. De exemplu, păreau să se bucure de anumite jucării, cântece sau videoclipuri şi chiar ajungeau să aştepte prezenţa regulată a unor asemenea lucruri în programul lor zilnic. Deşi comportamentul varia de la un copil la altul, unii dintre ei manifestau în mod clar iniţiativă (în limitele dizabilităţilor lor motorii) – de exemplu, lovind obiecte care făceau zgomot agăţate într-un cadru special construit în acest scop sau activând jucăriile preferate utilizând un întrerupător. Astfel de comportamente erau însoţite de semne de plăcere sau de entuziasm adecvate situaţiei.

E clar că aceşti copii nu pot fi descrişi drept vegetativi. Toate acestea sunt cu atât mai uimitoare cu cât copiii cu hidroanencefalie *se nasc fără cortex.* De obicei, cauza este un accident vascular masiv în uter, ce are drept consecinţă reabsorbţia prozencefalului, astfel încât craniul fătului se umple cu lichid cefalorahidian în

loc de ţesut cerebral. De aici şi termenul „hidroanencefalie”, care înseamnă „apă în loc de encefal”. O ilustrare a acestei afecţiuni găsiţi în figura 4: un RMN cerebral al unei fetiţe de trei ani născute fără cortex cerebral. Fotografiile înfăţişând un asemenea copil (figura 5) demonstrează reacţia ei emoţională când frăţiorul său îi este aşezat în poală.

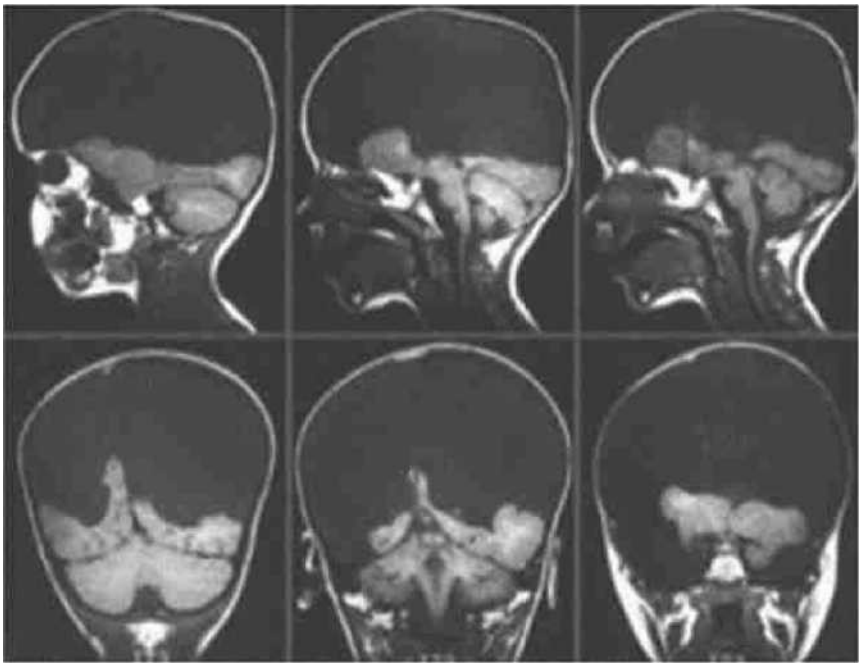


Figura 4. RMN cerebral al unei fetiţe de trei ani născute fără
cortex cerebral. Regiunea întunecată extinsă din interiorul
craniului indică lipsa ţesutului.

Figura 5. Reacţia unei fetiţe cu hidroanencefalie când frăţiorul
ei îi este aşezat în poală

Unde este generată conştiinţa în creier? De 150 de

ani, răspunsul aproape universal la această întrebare este „în cortex”. Aceasta a fost singura idee asupra căreia Freud şi curentul tradiţional principal din ştiinţa minţii din secolul XX au căzut de acord. Şi totuşi, dacă această afirmaţie este valabilă, atunci, când cortexul lipseşte, conştiinţa ar trebui să dispară. În cazul copiilor cu hidroanencefalie, acest lucru nu pare să se întâmple. Toate dovezile comportamentale sugerează că ei sunt de fapt conştienţi. Nu sunt comatoşi şi nici nu trăiesc într-o stare vegetativă.

Oare trebuie respinsă teoria conform căreia conştiinţa ar fi localizată în cortex? Propun să nu ne grăbim să tragem această concluzie. O obiecţie posibilă este că, la aceşti copii, cortexul nu a fost *îndepărtat* chirurgical la propriu. Procedura respectivă se numeşte „decorticare”.

Ce se întâmplă în cazul acestei intervenţii? Evident, asemenea proceduri experimentale nu pot fi efectuate pe copii, dar s-au realizat adesea la nou-născuţi ai altor mamifere, precum câini, pisici şi şobolani. Rezultatul este întotdeauna acelaşi: conform criteriilor comportamentale obiective pe care le folosim în mod obişnuit ca să o măsurăm, conştiinţa *se menţine*[*1]. Comportamentul postoperator al acestor animale nu poate fi descris drept „vegetativ” sau „comatos” oricât am lărgi definiţiile acestor termeni. Merker scrie că acestea nu prezintă „anomalii comportamentale majore ce i-ar permite unui observator obişnuit să le identifice ca având dizabilităţi”. Antonio Damasio e de acord: „Mamiferele decorticate prezintă o persistenţă

remarcabilă a comportamentului coerent, orientat spre obiective corespunzător senzaţiilor şi conştiinţei”[3]. De pildă, şobolanii supuşi procedurii de decorticare după naştere stau pe patru membre, se ridică, se caţără, se agaţă de bare şi dorm în poziţii normale. Se curăţă, se joacă, înoată, mănâncă şi se apără. Indiferent de sex, asemenea şobolani sunt capabili să se împerecheze cu succes cu parteneri normali ţinuţi în cuşcă. Când cresc, femelele manifestă elementele esenţiale ale comportamentului matern, care, deşi deficitar în anumite privinţe, le permite să-şi crească puii până la maturitate[4].

Situaţia este şi mai ciudată decât pare iniţial. Din multe puncte de vedere, mamiferele supuse decorticării cerebrale sunt de fapt *mai* active, *mai* emoţionale şi *mai* receptive decât cele normale. Panksepp obişnuia să le ceară studenţilor să aleagă între două grupuri de şobolani, determinând în funcţie de comportament grupul ai cărui membri fuseseră supuşi intervenţiei. De obicei, studenţii îi alegeau pe cei normali, gândind că şobolanii din celălalt grup (în cazul căruia se făcuse decorticarea) erau „mai vioi”[5].

Dacă susţinem ipoteza conform căreia conştiinţa ar fi localizată în cortex, atunci aceste animale vioaie – şi copiii expresivi, cu reacţii emoţionale observaţi de Merker la Disney World – trebuie să fie într-un anumit sens inconştiente. Cum ar fi posibil acest lucru?

Există un răspuns convenţional la această întrebare. Iată povestea. Într-un anumit sens, există doi creieri, care îndeplinesc aceleaşi funcţii şi care comunică între ei în anumite momente, dar care, altfel, nu au deloc

acelaşi statut. Unul (cortexul) este psihologic şi conştient. Celălalt (trunchiul cerebral) nu este deloc aşa. Informaţiile de la organele de simţ sunt transmise nu numai cortexului, ci şi coliculilor superiori ai trunchiului cerebral, printr-un set de conexiuni subcorticale (vezi figura 6). Aceste conexiuni prelucrează informaţiile

senzoriale, dar o fac în mod *inconştient.* Să luăm exemplul celebrului fenomen al „vederii oarbe”, ce apare atunci când cortexul vizual este distrus[6]. Pacienţii cu această afecţiune sunt capabili să răspundă la stimuli vizuali, dar, când li se cere să descrie ce „văd”, afirmă că nu au deloc imagini vizuale. În schimb, dacă li se cere, îşi folosesc instinctele pentru a *ghici* unde se află stimulii vizuali şi fac acest lucru cu o precizie remarcabilă. Să luăm cazul pacientului T.N., descris de specialistul în neuroştiinţe Lawrence Weiskrantz: deşi complet orb – cu alte cuvinte, cu totul lipsit de experienţe vizuale conştiente –, T.N. s-a strecurat cu iscusinţă printre obstacolele amplasate în calea sa de-a lungul unui coridor. Apoi, când a fost întrebat, a spus că habar n-avea că evita ceva[7]. Această capacitate surprinzătoare este posibilă deoarece calea de la nervul optic la coliculii superiori din trunchiul cerebral a rămas intactă, în pofida absenţei cortexului occipital[8].

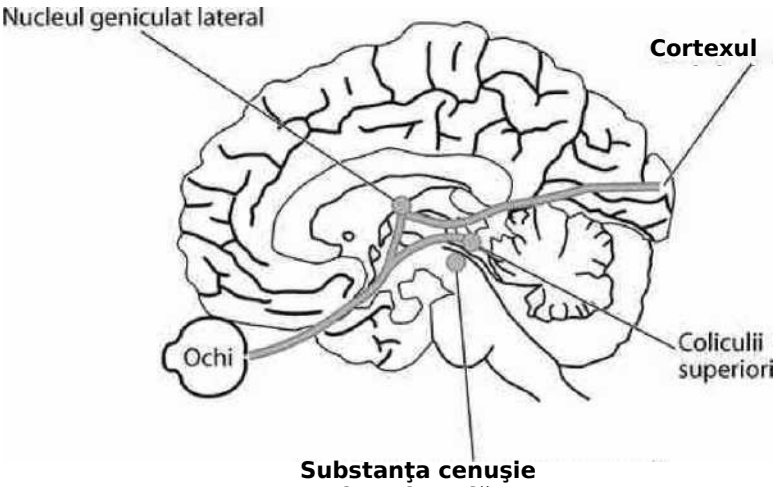


Figura 6. Această figură arată amplasarea coliculilor superiori

şi a cortexului occipital şi conexiunile lor cu ochiul. Conexiuni similare există şi pentru celelalte modalităţi senzoriale. Este indicată şi substanţa cenuşie periapeductală (PAG), despre care nu am discutat încă, dar a cărei relevanţă va fi evidentă în curând.

Se consideră că existenţa vederii sugerează că conştientizarea percepţiei vizuale survine probabil în cortex, şi nu în trunchiul cerebral. Se presupune că „nu e nimeni acasă” în trunchiul cerebral: acesta este o maşinărie autonomă ce prelucrează informaţiile vizuale în acelaşi mod inconştient în care o face o cameră de luat vederi. Acest principiu se aplică şi în cazul celorlalte simţuri, fiecare implicând propria zonă de conştientizare în cortex, dar fiecare (cu excepţia simţului olfactiv) transmiţând şi informaţii coliculilor superiori inconştienţi ai trunchiului cerebral.

Fetiţa înfăţişată în figura 5 nu are cortex funcţional. Dacă raţionamentul de mai sus este corect, probabil că ea îşi simte *inconştient* frăţiorul aşezat în poală, fără generarea unei percepţii conştiente a situaţiei. De fapt, ea trebuie să fie incapabilă de orice fel de experienţă conştientă. Ea ar fi cam ceea ce filosoful David Chalmers numeşte „zombi”. Deşi din anumite puncte de vedere ea acţionează normal, înăuntru e întuneric*[2].

Într-o analiză amplă a tipurilor a ceea ce numim „conştiinţă”, neurologul Adam Zeman face distincţie

între două sensuri principale ale termenului: „conştiinţa ca stare de veghe” şi „conştiinţa ca experienţă”[9]. Ulterior, Anton Coenen a elaborat: „Conştiinţa în primul ei sens (conştiinţa ca stare de veghe) este, din această perspectivă, o *condiţie necesară* pentru conştiinţa în al doilea sens (conştiinţa ca experienţă sau conştiinţa fenomenologică)”[10].

Cele două semnificaţii coincid cu distincţia convenţională în neurologie dintre „nivelul” cantitativ şi „conţinutul” calitativ ale conştiinţei. Prin urmare, e posibil ca, deşi animalele supuse procedurii de decorticare şi copiii cu hidroanencefalie sunt în stare de veghe, experienţa lor să fie lipsită de conţinut. Am văzut mai sus că sunt receptivi şi prezintă iniţiative comportamentale. Cu toate acestea, putem păstra ipoteza conform căreia cortexul este sediul „conştiinţei ca experienţă” dacă adăugăm că *a fi* conştient în sensul comportamental de a fi în stare de veghe şi a reacţiona este semnificativ diferit de *a avea* conştiinţă în sens fenomenologic – adică a fi subiect al experienţei.

În cazul în care noi doi semănăm, în acest punct poate că te simţi uşor neliniştit. Fata din figura 5 este conştientă în sensul că este trează şi are reacţii şi iniţiază singură acţiuni orientate spre obiective. În acelaşi timp, dacă „conştiinţă” înseamnă să ai experienţe fenomenologice, atunci probabil că ea nu are aşa ceva. Ca să folosesc o altă expresie populară în rândul filosofilor, *nu există „ceva cum e” să fii ea.*

Iată ce mi se pare atât de tulburător în privinţa acestui raţionament, chiar dacă sună puţin naiv.

Analizându-mi propria situaţie, a avea reacţii şi a fi în stare de veghe şi a avea o experienţă conştientă sunt, mai mult sau mai puţin, acelaşi lucru. Din câte ştiu, eu nu sunt niciodată în stare de veghe şi am reacţii, dar sunt şi inconştient din punct de vedere fenomenologic. Aceste două aspecte merg mână-n mână. De îndată ce mă trezesc, devin conştient de diverse lucruri. De fapt, în opinia mea, mi se pare că conştiinţa mea interioară îmi *cauzează* reacţia la stimuli externi, cel puţin într-o anumită măsură. De regulă, când observ ceva – adică atunci când conştientizez acest ceva –, am şi o reacţie din punct de vedere intenţional. Presupun că şi în cazul tău e la fel.

Acesta este motivul pentru care ne facem o idee cu privire la conştiinţa pacienţilor cu antecedente neurologice în funcţie de reacţiile lor. Ce altceva mai putem face? În practica clinică, facem distincţie între stări precum coma, starea vegetativă şi starea de veghe cu reacţii complete cu ajutorul scalei de comă Glasgow în 15 puncte. Această scală este alcătuită din teste ale răspunsului ocular al pacienţilor, ale răspunsurilor lor verbale la întrebări şi ale răspunsului lor motor la instrucţiuni şi (dacă este necesar) la durere[11]. Dacă pacientul are răspunsuri complete, îl considerăm conştient şi îl tratăm ca atare. Nu ne punem problema că ar putea răspunde *ca şi cum* ar fi conştient când, de fapt, este zombi. Problema apare în sens invers: neurologii trebuie să verifice dacă există posibilitatea ca pacientul să nu aibă reacţii la nivel exterior, dar să fie conştient la nivel interior – cum se întâmplă, de pildă, în cazuri cu pierderea totală a motricităţii precum

pseudocoma.

Ce putem conchide de aici? E de netăgăduit că problema filosofică a altor minţi ridică semne de întrebare cu privire la ce putem şti despre stările subiective ale animalelor şi ale altor oameni din comportamentul lor, cam tot aşa cum minciunile sau prefăcătoriile ne pot induce în eroare cu privire la stările interioare ale unei persoane. Dar problema altor minţi *nu* stabileşte că experienţa interioară şi reacţia exterioară sunt de fapt independente una de alta. Întrucât fiecare simte prezenţa unei legături puternice în cazul său individual, sarcina probei le revine cu siguranţă celor care vor să pretindă că comportamentul de răspuns emoţional nu implică experienţa fenomenologică. Iată ce prevăd în mod normal regulile ştiinţei.

Aceste afirmaţii sunt valabile mai ales când avem de-a face cu fiinţe umane despre care, indiferent de ce afecţiuni neurologice ar mai suferi, putem presupune că îşi generează comportamentul din starea de veghe într-un mod oarecum obişnuit – iar prin asta mă refer la funcţionarea normală a regiunilor cerebrale fără leziuni, şi nu la minciuni şi prefăcătorie şi nici la tipul de născociri viclene pe care ne-am putea aştepta să le găsim la nişte roboţi filosofici imaginari.

Implicaţiile pentru etica medicală sunt considerabile. O colegă de-a mea, psihiatră, al cărei copil are hidroanencefalie mi-a mărturisit recent o dilemă teribilă cu care s-a confruntat după ce a devenit mamă. Neurochirurgul care îi trata copilul i-a sugerat că o intervenţie pentru închiderea fontanelei se putea face

fără anestezie deoarece, neavând cortex, copilul era, evident, incapabil să simtă durere. Colega mea nu mi-a spus ce decizie a luat, iar întrebarea ar fi fost prea delicată. Totuşi, putem vedea din asemenea cazuri cum unele consideraţii teoretice aparent abstracte pot duce foarte uşor la erori medicale îngrozitoare. Prin urmare, vreau să subliniez foarte clar: ca să acceptăm că o persoană ce *pare* să fie conştientă *nu este* de fapt, trebuie să cerem argumente extrem de convingătoare. Simpla invocare a îndoielii filosofice nu este suficientă. Avem nevoie de argumente foarte puternice ca să credem că cele două tipuri de conştiinţă s-au separat la asemenea persoane, întrucât se pare că nouă nu ni se întâmplă niciodată aşa ceva.

În prezent, aceste argumente sunt furnizate de teoria conform căreia conştiinţa survine exclusiv în cortex. E ceva atât de obişnuit să se accepte această idee în domeniul neuroştiinţelor behavioriste, încât nu mi-a trecut prin cap să o pun la îndoială până când, în baza cercetărilor mele despre vise, Allen Braun mi-a atras atenţia asupra rolului misterios jucat de trunchiul cerebral. În virtutea teoriei conform căreia conştiinţa survine exclusiv în cortex, doctorii hotărăsc dacă îşi tratează sau nu pacienţii ca pe nişte fiinţe capabile de simţire. De exemplu, unii îngrijitori se ocupă de copii cu hidroanencefalie în condiţii de neglijare emoţională severă în baza supoziţiei că cei mici sunt „vegetativi”. Dacă teoria este corectă, atunci nu avem nicio problemă: nu e vorba de fapt de neglijare deoarece copiii sunt asemenea zombilor filosofici. Ei par să aibă sentimente, dar de fapt nu au. Impresia că au este o

iluzie produsă de problema altor minţi. Lăsându-ne înşelaţi de comportamentul lor exterior, ne imaginăm că au o oarecare natură interioară[12].

Aşadar, ce este teoria conform căreia conştiinţa survine exclusiv în cortex şi cât e de convingătoare?

Primul lucru de observat este că ea a început să se dezvolte foarte devreme. Observaţia curentă conform căreia conştiinţa noastră constă în principal din imagini perceptive ale unor evenimente ce se desfăşoară în jurul nostru sugerează că fluxul conştiinţei spre interior trece prin simţuri. Fără îndoială, această opinie de bun-simţ s-a perpetuat de când am început să ne gândim la asemenea chestiuni. În secolele al XVII-lea şi al XVIII-lea ea a dat naştere filosofiilor „empiriste” ale lui John Locke şi David Hume. Ei au formulat teoria că mintea – care la început e o tabula rasa – îşi capătă toate caracteristicile specifice prin impresiile lăsate de vibraţiile senzoriale. Se presupunea că impresiile se asociază unele cu altele prin îmbinări regulate de diferite tipuri[13] pentru a ne produce *imaginile mnezice* ale obiectelor care, la rândul lor, devin pietrele de temelie ale unor idei mai abstracte. Vibraţiile senzoriale ulterioare stimulează aceste imagini asamblate în prim-planul conştiinţei, astfel încât ceea ce trăim nu sunt senzaţii brute, ci *ceea ce am învăţat* despre lume.

Modul în care ideile devin conştiente ca răspuns la un stimul extern a fost numit *apercepţie* (care înseamnă aproximativ percepţia prezentului prin prisma experienţelor trecute)[14]. S-a spus că procesele cognitive

precum imaginile mentale folosite în gândire implică aproximativ acelaşi proces în sens invers: activări generate intern ale imaginilor mnezice, rearanjate adecvat (şi mai slabe decât cele generate extern).

Deşi era speculativă, această idee filosofică despre minte a devenit harta după care s-au ghidat primii neurologi. Când pionierii din secolul al XlX-lea ai neuropsihologiei moderne au căutat să stabilească corelativele neurale ale acestor procese, au observat că organele de simţ erau conectate la cortex şi au presupus că „vibraţiile" senzoriale aveau loc în aceşti nervi de legătură. Se întâmplă că nu au trecut cu vederea faptul că organele de simţ erau conectate şi cu nucleii subcorticali. Totuşi, au presupus că marele depozit de imagini mnezice care constituie cunoştinţele noastre despre lume se află în cortex, deoarece acesta conţine incomparabil mai mulţi neuroni. Astfel, s-a presupus că „apercepţia" şi „ideile" aferente care generează activitatea mintală în sine sunt fenomene corticale. Theodor Meynert, marea autoritate a epocii în materie de neuroanatomie, s-a exprimat astfel:

> Principala funcţie a organului central este să transmită faptul existenţei către un eu care se modelează treptat în fluxul creierului [...]. Dacă privim cortexul drept un organ ce funcţionează ca un întreg, atunci se poate spune numai că el este util în procesele minţii [...]. E imposibil şi nenecesar să analizăm cortexul mai mult de-atât[15].

Cortexul este, desigur, impresionant de mare la

oameni[16]. Neurologul Alfred Walter Campbell, specialist în neuroanatomie, a rezumat acest curent de gândire dominant în cadrul întâlnirii anuale a Asociaţiei de Medicină şi Psihologie din Londra, în 1904:

> Văzut în ansamblu, creierul uman include două tipuri de centri, controlând ceea ce putem numi funcţii „evoluţionare primare” şi, respectiv, „superioare”; primele sunt cele comune la toate animalele şi esenţiale pentru supravieţuire, adică centrii pentru mişcare şi senzaţii comune şi speciale; cele din urmă sunt funcţiile psihice complexe care îl fac pe om să se ridice deasupra tuturor celorlalte fiinţe[17].

Este important să recunoaştem că, din această perspectivă, mintea constă în întregime din imagini mnezice reflectând experienţele trecute ale lumii exterioare. Senzaţiile noi pur şi simplu stimulează aceste imagini şi asocierile lor în conştiinţă. Prin urmare, vibraţiile ce intră prin intermediul organelor de simţ sunt evenimente prepsihice – *declanşatori* ai activităţii mintale –, ele nefiind evenimente psihice. Aceeaşi afirmaţie este valabilă şi în cazul funcţiei nervilor ce conectează cortexul cu restul corpului: căile lor nu au nimic „psihic”, ele pur şi simplu eliberează *outputurile* activităţii psihice. Activitatea mintală propriu-zisă nu poate avea loc decât în cortex, unde se află imaginile mnezice.

Meynert a descris relaţia dintre cortex şi lumea exterioară după cum urmează:

> Efectele motoare ale conştiinţei noastre ce răspunde la lumea exterioară nu sunt rezultatul unor forţe inerente creierului. Creierul, ca o stea fixă, nu radiază căldură: îşi obţine energia susţinând toate fenomenele cerebrale din lumea de dincolo de el[18].

Probabil merită să evidenţiem şi că, întrucât Freud încă nu apăruse, se presupunea că toată activitatea psihică este conştientă. Se considera că „psihic” şi „conştient” înseamnă acelaşi lucru.

Activităţile experimentale din secolul al XlX-lea care au condus la reprezentarea modernă a conştiinţei s-au desfăşurat exclusiv în acest cadru filosofic. La sfârşitul anilor 1800, fiziologul Hermann Munk a identificat cortexul occipital drept sediul aspectului mental al vederii (vezi figura 6). A devenit interesat de comportamentul câinilor cu leziuni experimentale la nivelul cortexului occipital. Aceste nefericite animale puteau să vadă, dar se pare că le lipsea „înţelegerea” normală a ceea ce vedeau: de exemplu, nu mai puteau să îşi recunoască vizual stăpânii sau să-şi identifice castronul cu mâncare, deşi se uitau la ele şi le dădeau ocol, dar le puteau recunoaşte cu ajutorul celorlalte simţuri[19]. Munk a numit această afecţiune „cecitate psihică”, pentru a o deosebi de forma comună de orbire cauzată de leziuni la nivelul căilor senzoriale (subcorticale) ducând de la ochi la cortex[20]. În spiritul filosofiei empiriste, a pus semnul egalităţii între ceea ce numea vedere „psihică” şi capacitatea de a activa *imaginile mnezice* vizuale prin asocieri, în opoziţie cu procesul mecanic al receptării senzaţiilor vizuale brute

şi al declanşării unor reflexe motoare. În consecinţă, afecţiunea descrisă de Munk este numită acum „agnozie” vizuală – adică lipsa *cunoaşterii* vizuale.

Fenomene clinice de exact acelaşi tip au fost curând raportate şi în cazul unor oameni – de exemplu, în 1887, de oftalmologul Hermann Wilbrand, care a descris cazul lui Frăulein G., o femeie de 63 de ani care a suferit un accident vascular occipital bilateral[21]:

> Toţi din jur o considerau oarbă. Totuşi, ea era conştientă că nu era complet oarbă, „pentru că, atunci când oamenii stăteau la căpătâiul meu şi vorbeau cu milă de orbirea mea, îmi spuneam: nu pot fi cu totul oarbă pentru că pot să văd faţa de masă de acolo, cu margine albastră, întinsă pe masa din infirmerie” [...]. Când această doamnă – altfel foarte inteligentă – s-a ridicat [după ce iniţial şi-a pierdut cunoştinţa, din cauza accidentului vascular], şi-a dat seama că era într-o stare ciudată în care nu vedea, fiind totuşi capabilă să vadă [...]. Şi astăzi este tulburată când îşi aminteşte prima ieşire din casă după accidentul vascular; cât de diferit şi de ciudat i se părea oraşul şi cât de tulburată şi de angoasată s-a simţit când a fost condusă de însoţitoarea ei pentru prima dată pe Jungfernstieg şi de la Neuer Wall la Stadthaus şi cum, după indicaţiile însoţitoarei, clădirile şi străzile care îi erau de obicei atât de familiare parcă erau ceva nou. Iată cum îşi relatează Frăulein G. reacţia la afirmaţiile femeii care o însoţea la vremea respectivă: „Dacă spui că aici este Jungfernstieg, acolo Neuer Wall, iar acolo Stadthaus,

presupun că aşa trebuie să fie, deşi eu nu le recunosc [...]. I-am spus medicului meu: «Din afecţiunea mea se poate trage concluzia că vedem mai mult cu creierul decât cu ochii; ochiul nu este decât vehiculul vederii, deoarece eu văd orice lucru în mod absolut clar şi lucid, dar nu recunosc lucrul şi adesea nu ştiu ce ar putea fi lucrul pe care îl văd»".

Wilbrand a concluzionat că această pacientă suferise o pierdere a memoriei vizuale (adică o tulburare a recunoaşterii, înţelegerii sau cunoaşterii vizuale), şi nu o simplă orbire. Şi a observat că, deşi pacienţii orbi încă pot genera imagini vizuale în visele lor, întrucât capacitatea lor de a produse imagini mentale este intactă, persoanele oarbe *psihic* precum Fräulein G. nu pot; ele îşi pierd capacitatea de a visa. Şi s-a întrebat cum poţi genera halucinaţii vizuale fără imagini mnezice vizuale[22].

Asemenea observaţii despre vedere erau generalizate pe atunci şi la alte modalităţi de percepţie. Astfel, ablaţia cortexului auditiv (din nou, la câini) producea „surzenie mintală", numită astăzi agnozie auditivă, animalele afectate nemaiputând răspunde la sunete cu semnificaţie dobândită, chiar dacă în mod clar nu erau surde: răspundeau la zgomote brute, dar nu îşi mai recunoşteau numele când erau chemate. În 1874 neurologul Carl Wernicke a observat ceva similar la pacienţii umani, iar astfel a formulat ideea de tulburări ale limbajului dobândit numite „afazie"[23].

La nivel motor, Hugo Liepmann a afirmat că şi leziunile corticale produc deficienţe ale aspectului

psihic al mişcării, numite „paralizie psihică” (sau „apraxie”). O leziune la nivelul căilor motoare de ieşire a cauzat paralizie fizică, dar o leziune la nivelul centrului cortical pentru imaginile mnezice motoare a cauzat uitarea abilităţilor de mişcare dobândite (numită apraxie „chinestezică a membrelor”), iar o leziune la nivelul căilor transcorticale pentru asocierile motoare a cauzat o întrerupere între mişcarea deprinsă şi ideile abstracte, cum ar fi semnificaţia simbolică a gesturilor mâinii (apraxie „ideomotoare”)—.

Orbirea, surditatea şi paralizia subcorticale au fost astfel conceptualizate ca afecţiuni *fizice,* iar efectele vizuale, auditive şi motoare ale leziunilor corticale erau *mintale:* formele aperceptive şi asociative de agnozie, afazie şi apraxie. Aceste din urmă afecţiuni sunt numite astăzi tulburări „neurocognitive”. Astfel, distingerea între aceste două clase de tulburări (subcorticale versus corticale) a coincis cu trasarea graniţei disciplinare dintre neurologie şi neuropsihologie şi rămâne valabilă.

Când contemporanii lui Munk au extirpat complet cortexul la câini – nu doar părţile specializate responsabile cu anumite tipuri senzorimotoare –, animalele nu au intrat în comă sau într-o stare vegetativă, ci s-au comportat ca şi cum ar fi fost *fără minte* (în accepţiunea empiristă a termenului). Animalele au devenit amnestice, adică şi-au pierdut toate imaginile mnezice şi, prin urmare, „înţelegerea”. În consecinţă, Friedrich Goltz le-a descris drept „nătânge”. Faptul că nu au intrat în comă nu i-a surprins pe primii cercetători deoarece, având în vedere ipotezele teoretice ale vremii, ei se aşteptau doar ca

aceşti câini supuşi procedurii de decorticare cerebrală să-şi piardă toate *cunoştinţele* dobândite. Nu era neapărat ca această pierdere să aibă efect şi asupra senzaţiilor şi reflexelor corporale, despre care se presupunea că sunt subcorticale. Ideea că viaţa psihică implică doar imagini mnezice nu era controversată. La vremea respectivă, controversele vizau alte chestiuni, cum ar fi cât de precis putea fi localizată fiecare funcţie psihică în regiuni specifice ale cortexului. Nimeni nu se îndoia că toate funcţiile de acest fel sunt corticale.

Aşa că, atunci când Meynert a rezumat tot tabloul ce se contura în celebra lui carte din 1884 *Psychiatry*[25], a identificat mintea cu *totalitatea imaginilor mnezice ale obiectelor* produse prin proiecţia periferiei senzorimotoare în cortex plus asocierile transcorticale dintre ele şi imaginile mnezice care constituiau ideile abstracte. În mod firesc, a localizat toate aceste imagini, asocieri şi idei în cortex – conform citatului de mai sus. L-a numit partea „voluntară” a creierului şi a pretins că are legături directe cu periferia trupului prin nervii senzoriali şi motori. Adică, în opinia sa, cortexul era conectat cu corpul *independent* de substanţa cenuşie subcorticală. A recunoscut că partea subcorticală a creierului era legată de asemenea de cortex şi de periferia corpului, prin propriile căi distincte, dar le-a descris drept „reflexe”. Aici avem cei „doi creieri” pe care i-am menţionat mai sus, care explică comportamentul copiilor cu hidroanencefalie.

În acest fel, cortexul a devenit organul minţii – mintea concepută drept *conştiinţă a imaginilor mnezice* –, iar creierul subcortical a devenit fără minte. Totul se

reducea la ideea că „mintea” este constituită în întregime din experienţe trecute, care lasă urme asociate între ele pentru a forma imagini concrete şi idei abstracte. Restul, părţile periferice senzorimotoare şi părţile înnăscute şi subcorticale, inclusiv părţile care transmit impresii din *interiorul* corpului, erau considerate exclusiv reflexe. Şi aşa, deşi pare ciudat, distincţia filosofică dintre minte şi corp a ajuns să coincidă cu distincţia anatomică dintre cortex şi subcortex.

În anii 1960, Norman Geschwind, marele pionier al neurologiei behavioriste din America, a reluat cu entuziasm aceste idei clasice[26]. Renaşterea asociaţionismului empirist în neurologie a coincis cu „revoluţia cognitivă” din psihologie. Schemele cu informaţii ale specialiştilor în ştiinţe cognitive semănau foarte mult cu diagramele folosite de specialiştii germani clasici în neurologie pentru a ilustra relaţiile funcţionale dintre centrii corticali ce conţin imagini mnezice de diferite tipuri. Dar – după cum vom vedea – cognitivismul modern a condus şi la conştientizarea (consacrată până în anii 1990) a faptului că multe funcţii psihice (inclusiv percepţia şi memoria) *nu sunt totuşi conştiente.*

Cam în perioada în care Geschwind reintegra în neuropsihologie teoria clasică potrivit căreia conştiinţa apare exclusiv în cortex, pionierii (precum Paul MacLean şi James Olds) a ceea ce avea să devină neuroştiinţa afectivă acumulau observaţii care au dezvăluit că funcţii psihice îndeplineau şi numeroşi nuclei subcorticali ce conectau creierul cu interiorul

corpului, funcţii precum *imboldul* motivaţional şi *trăirea* emoţională.

În pofida acestor evoluţii paralele în neuroştiinţele cognitive şi afective, pe care le voi elabora mai jos, până la urmă această moştenire intelectuală – încercarea neurologilor din secolul al XIX-lea de a confirma teoriile filosofilor din secolul al XVIII-lea – e cea pe care se clădeşte prezumţia neurologilor de astăzi conform căreia copiii cu hidroanencefalie sunt fără minte. Ei fac această prezumţie, deşi recunosc că echivalarea de altădată a „psihicului" cu conştiinţa şi cu cortexul nu mai este valabilă: că nu întreaga activitate psihică este conştientă şi nu toată conştiinţa este corticală. Cu alte cuvinte, insistenţa asupra ideii că această echivalare este încă valabilă, că „conştiinţa ca experienţă" e cu necesitate corticală are la bază o inerţie a teoriei, şi nu nişte dovezi ştiinţifice.

După cum poate ţi-ai dat seama, eu nu consider că teoria conform căreia conştiinţa survine exclusiv în cortex este valabilă. De fapt, aş spune chiar că animalele şi oamenii pot fi conştienţi chiar dacă nu au deloc cortex. Eu cred că *există* „ceva cum e" să fii un copil cu hidroanencefalie.

Desigur, unul dintre motivele pentru care e atât de dificil să ştim ce trăiri au pacienţii cu hidroanencefalie – dacă au vreo experienţă interioară – este că ei nu pot să vorbească. Limbajul chiar este o funcţie corticală, aşa că nu ne putem aştepta ca oamenii care nu o au să ne furnizeze rapoarte verbale introspective. Ei nu ne pot furniza tipul de dovadă subiectivă care de obicei ne face

să credem că alţi oameni sunt conştienţi, în pofida problemei altor minţi[27]. Afirmaţia e valabilă şi în cazul animalelor. Totuşi, există şi oameni care şi-au pierdut o porţiune substanţială din cortex, dar nu şi capacitatea de a vorbi. În aceste cazuri, pur şi simplu îi putem *întreba* cum stă treaba.

Neurochirurgii Wilder Penfield şi Herbert Jasper au extirpat porţiuni mari din cortex (chiar emisfere întregi) la 750 de pacienţi umani sub anestezie locală, în principal pentru tratarea epilepsiei. Au observat că asemenea intervenţii au efecte limitate asupra conştiinţei autoraportate, chiar în momentul în care cortexul este îndepărtat. (Intervenţiile chirurgicale pe creier se fac adesea sub anestezie locală, astfel încât pacientul să poată raporta efectele a ceea ce face chirurgul.) Penfield şi Jasper au concluzionat că rezecţiile corticale nu întrerup existenţa conştientă, ci doar îi privează pe pacienţi de „anumite forme de informare”[28]. Eu am asistat la multe asemenea intervenţii, în timpul cărora de obicei rolul meu e să evaluez efectele stimulării electrice a ariilor din cortex responsabile cu memoria şi limbajul înainte de intervenţia chirurgului. Am observat exact ceea ce au raportat Penfield şi Jasper.

Dar nu tot cortexul este la fel. Regiunile cortexului care primesc inputuri de la fiecare dintre simţurile noastre speciale generează „qualia” fenomenale asociate cu simţurile respective (culoare pentru vedere, ton pentru auz şi aşa mai departe). Totuşi, conform opiniei standard, senzaţiile brute pot fi declarate numai după ce au fost accesate de un tip general de conştiinţă

care constituie „eul” conştient[29].

Sunt utilizate diferite terminologii pentru a descrie relaţia dintre conştiinţa „fenomenală” (de exemplu, simpla vedere şi auzire) şi conştiinţa „reflexivă” *(ştii* că vezi şi auzi). Dar toate transmit aceeaşi idee de bază: eşti mai mult decât diversele forme de informaţii senzoriale pe care le prelucrezi. Aşa se explică faptul că, deşi unii pacienţi cu leziuni la nivelul cortexului vizual primar sunt orbi, iar unii cu leziuni la nivelul cortexului auditiv sunt surzi, ei nu şi-au pierdut simţul *sinelui.* Acest sine conştient ghiceşte ce vezi când nu ai qualia vizuale în cazul vederii oarbe, de exemplu. Asemenea pacienţi pierd doar „anumite forme de informare”. Aşadar, întrebarea arzătoare este: ce învăţăm de la pacienţii care şi-au pierdut porţiunile de cortex despre care se spune că sunt responsabile de ipseitate?

Există trei locuri ce candidează pentru această funcţie. Primul este *cortexul insular,* care este specializat în conştientizarea interoceptivă şi despre care se consideră în general că generează sentimentele ce constituie „sinele” conştient[30]. A doua este *cortexul prefrontal dorsolateral,* care formează o suprastructură peste toate celelalte părţi ale creierului şi despre care se consideră în general că permite „gânduri de ordin superior”, inclusiv conştientizarea sentimentelor[31]. A treia este *cortexul cingular anterior,* care se activează în mai toate experimentele de imagistică cognitivă a creierului şi despre care se spune că mediază aspecte precum „prelucrarea informaţiilor despre sine” şi „voinţa”[32].

Să luăm aceste trei regiuni pe rând.

Iată un fragment dintr-un interviu luat de Antonio Damasio unui pacient al cărui cortex insular a fost distrus complet de o infecţie virală numită encefalită cu *herpes simplex*[33]*:*

Întrebare: Ai un simţ al sinelui?
Răspuns: Da, am.
Î: Dar dacă ţi-aş spune că nu eşti aici acum?
R: Aş spune că eşti orb sau surd.
Î: Crezi că alţi oameni îţi pot controla gândurile?
R: Nu.
Î: Şi de ce crezi că nu e posibil?
R: Îţi controlezi singur mintea, să sperăm.
Î: Dacă ţi-aş spune că mintea ta este mintea altuia?
R: Când a avut loc transplantul, adică transplantul de creier? *Î*: Dar dacă ţi-aş spune că te cunosc mai bine decât te cunoşti tu pe tine?
R: Aş crede că te înşeli.
Î: Dar dacă ţi-aş spune că eşti conştient că eu sunt conştient? *R*: Aş spune că ai dreptate.
Î: Eşti conştient că eu sunt conştient?
R: Sunt conştient că eşti conştient că sunt conştient.

Acest pacient, numit „B.”, a fost studiat de echipa lui Damasio o perioadă de 27 de ani (s-a îmbolnăvit la 48 de ani şi a murit la 73 de ani). În studiul de caz publicat, echipa a raportat date neuropsihologice detaliate referitoare la capacitatea lui de a simţi şi de a răspunde emoţional. Iată concluziile principale:

Pacientul B., al cărui cortex insular a fost distrus în întregime, avea senzaţii corporale şi trăiri emoţionale. Raporta că simţea durere, plăcere, mâncărime, gâdilituri, fericire, tristeţe, frică, iritare, afecţiune şi compasiune şi se comporta în moduri în acord cu ceea ce simţea în momentele respective. Raporta şi senzaţii de foame, de sete şi dorinţa de a excreta şi avea comportamentele aferente. Tânjea să se joace – de exemplu, să joace dame –, să-i viziteze pe alţii, să se plimbe şi manifesta o plăcere evidentă când se implica în asemenea activităţi, precum şi dezamăgire şi chiar iritare când nu mai avea acces la astfel de oportunităţi. [...] Dată fiind sărăcirea imaginaţiei sale, existenţa pacientului B. era o reacţie „afectivă" practic continuă la propriile stări fizice şi la cerinţele modeste ale lumii din jurul său, care nu erau moderate de mecanisme de control cognitiv de ordin superior [...]. Conform lui Craig (2011), semnul ce confirmă conştiinţa de sine este capacitatea de a te recunoaşte când te vezi într-o oglindă, capacitate care, în cuvintele sale, „nu poate fi furnizată decât de o reprezentare neuronală a sinelui funcţională, validă din punct de vedere emoţional". Pacientul B. a trecut acest test în mod constant şi repetat. Pe scurt, aceste concluzii contrazic ipoteza conform căreia conştiinţa de sine la oameni, alături de capacitatea de a simţi, ar depinde în întregime de cortexul insular şi, în mod specific, de treimea sa anterioară (Craig, 2009, 2011). În lipsa cortexului insular, trebuie să luăm în considerare alternative neuroanatomice ca să explicăm

fundamentul capacităţilor senzoriale şi al facultăţii lui B. de a simţi[34].

Bud Craig şi alţii afirmă că trăirile afective ce constituie sinele sunt generate în cortexul insular. Şi totuşi, Damasio a descoperit că, de fapt, pacientul B. era, ca şi şobolanii supuşi procedurii de decorticare, *mai* emoţional după ce a suferit leziunile la nivelul cortexului decât înainte. Predicţia că asemenea pacienţi ar trebui să-şi piardă „prezenţa” conştientă este în mod clar invalidată.

Acelaşi lucru se întâmplă şi la pacienţii umani cu leziuni la nivelul celei de-a doua regiuni analizate aici: cortexul *prefrontal,* care constituie o suprastructură deasupra tuturor celorlalte părţi ale creierului. Impresionabilitatea pacienţilor cu leziuni la nivelul cortexului prefrontal este larg recunoscută, constituind o trăsătură esenţială în „sindromul de lob frontal”. Acest lucru fusese deja observat în 1868, în celebrul caz a lui Phineas Gage, după ce o bară de fier îi străpunsese craniul într- un accident de muncă:

> Echilibrul, ca să spunem aşa, dintre facultăţile sale intelectuale şi înclinaţiile lui animalice pare să fi fost distrus. Este schimbător, nepoliticos, dedându-se uneori la cele mai ordinare obscenităţi (ceea ce nu îi stătea în obicei înainte), manifestând prea puţin respect faţă de semenii lui, respingând cumpătarea sau sfaturile când intră în conflict cu dorinţele sale[35].

Unii dintre principalii specialişti în neuroştiinţe care

studiază emoţiile, precum Joseph LeDoux, consideră că sentimentele *se ivesc* în sensul propriu în cortexul prefrontal dorsolateral[36]. Conform acestei opinii, precursorii subcorticali ai sentimentelor sunt în întregime inconştienţi până când sunt „etichetaţi” în memoria funcţională conştientă[37]. Pentru aceşti teoreticieni, emoţia este doar o altă formă de cogniţie – de fapt, o formă de cogniţie reflexivă, mai degrabă abstractă. Dar, dacă au dreptate, atunci de ce pacienţii a căror memorie funcţională este în mare parte distrusă manifestă atât de multă emoţie? Phineas Gage este doar cel mai celebru caz dintr-o literatură consistentă. Ce ar putea să adauge la toată această imagine conştiinţa reflexivă, dacă forma prereflexivă a simţirii este atât de viu exprimată în comportamentul acestor pacienţi?

Trebuie să recunoaştem că majoritatea susţinătorilor teoriei conform căreia conştiinţa survine exclusiv în cortex nu sunt interesaţi în mod special de sentimente. Ei se axează în schimb pe contribuţia lobilor prefrontali la *gândirea* de ordin superior. De asemenea, ipotezele ce rezultă din aceste teorii mai generale pot fi testate imediat în cazul unor oameni cu leziuni la nivelul lobilor prefrontali. Leziunile prefrontale totale sunt foarte rare – dar nu inexistente.

Pacientul meu „W.” avea 48 de ani când l-am examinat prima dată. A obţinut scorul Glasgow maxim, ceea ce înseamnă că era complet conştient şi avea răspunsuri complete. La vârsta de 13 ani suferise un anevrism cerebral rupt, ce a necesitat o intervenţie în care lobii frontali au fost retractaţi pentru a acoperi vasul de sânge vulnerabil şi, astfel, a preveni altă

hemoragie. Din păcate, această intervenţie a cauzat o infecţie cerebrală cronică, presupunând multiple intervenţii suplimentare ce au dus în cele din urmă la distrugerea completă a lobilor prefrontali de ambele părţi (vezi figura 7).

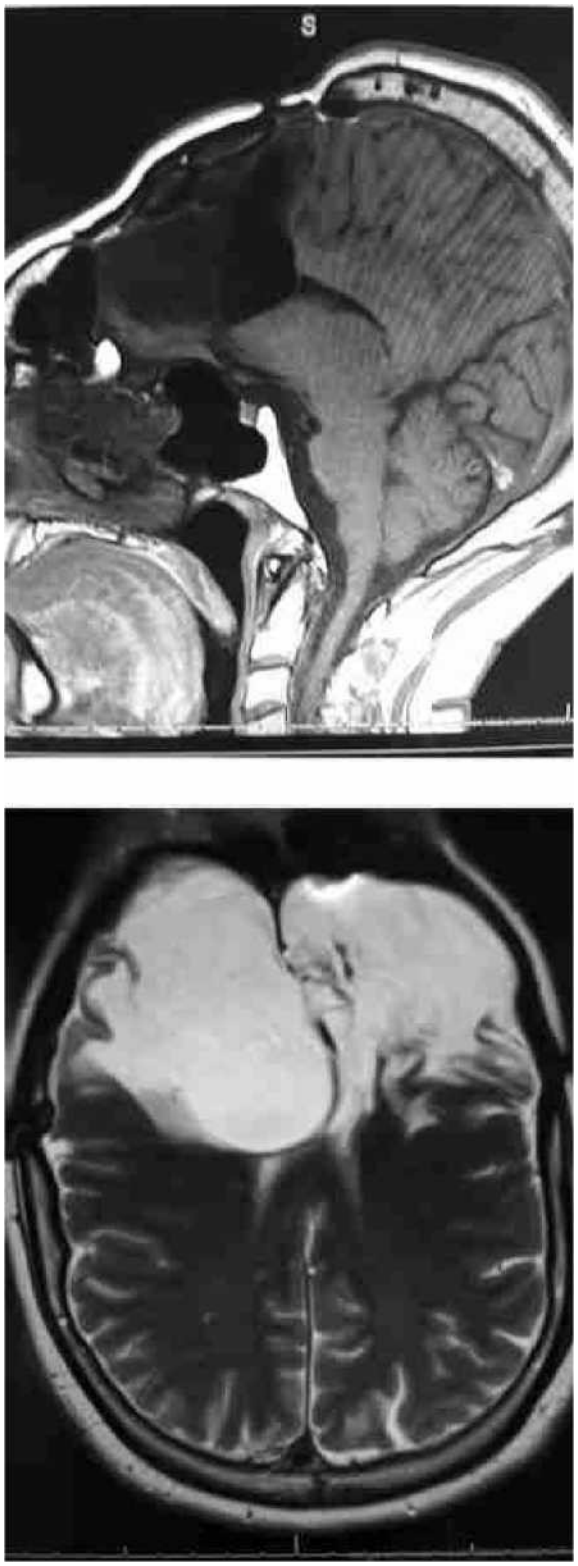


Figura 7. RMN cerebral al pacientului W., ce arată distrugerea

completă a lobilor prefrontali de ambele părţi

Din fericire, regiunea cerebrală responsabilă cu limbajul nu a fost afectată. Iată o parte din conversaţia noastră:

Întrebare: Eşti conştient de gândurile tale?
Răspuns: Da, sigur că da.
Î: Pentru a confirma acest lucru, te rog să rezolvi o problemă în care ţi se cere să-ţi imaginezi în mod conştient o situaţie în minte.
R: OK.
Î: Imaginează-ţi că ai doi câini şi un pui de găină.
R: OK.
Î: Îi vezi cu ochii minţii?
R: Da.
Î: Acum spune-mi câte picioare vezi în total?
R: Opt.
Î: Opt?
R: Da, câinii au mâncat puiul.

Poanta a fost spusă cu un rânjet răutăcios. Poate că nu e cea mai bună glumă din lume, dar constituie o dovadă convingătoare că, folosind o expresie colocvială, era cineva acasă. Pacientul W. a afirmat că este conştient, cum a afirmat şi pacientul B. al lui Damasio, iar în lipsa unui argument mai convingător decât scepticismul filosofic radical, eu înclin să-i cred pe cuvânt.

A treia şi ultima regiune corticală despre care se presupune că are o relaţie specială cu „eul” conştient este *girusul cingular anterior*. Pacienţii cu leziuni

bilaterale totale în această regiune sunt relativ des întâlniţi, parţial din cauză că această regiune a cortexului este o ţintă obişnuită a intervenţiilor chirurgicale în

afecţiunile psihiatrice, precum tulburarea obsesiv-compulsivă. (Regiunea cingulară anterioară şi-a atras toată această atenţie pe cap, cum s-ar spune, tocmai din cauza asocierii sale cu „prelucrarea informaţiilor legate de sine”).

În perioada postoperatorie acută, unii dintre aceşti pacienţi nu mai pot face diferenţa între fantezie şi realitate. Charles Whitty şi Walpole Lewin au descris un exemplu frapant (Cazul 1, în lucrarea lor):

> Ca răspuns la întrebarea referitoare la ce a făcut în ziua respectivă, pacientul a spus: „Eu am stat la ceai cu soţia mea”. Fără alte comentarii din partea intervievatorului, a continuat: „A, de fapt eu n-am stat la ceai. Ea n-a fost azi pe- aici. Dar, imediat ce eu închid ochii, scena asta îmi apare într-un mod atât de viu. Eu văd ceştile şi farfurioarele şi o [illegible] aud cum toarnă ceaiul. Apoi, cum eu ridic ceaşca să iau o gură, eu mă trezesc şi nu mai e nimeni”.
> Întrebare: Deci dormi mult?
> Răspuns: Nu, eu nu dorm – e un fel de vis în stare de veghe [...], chiar cu ochii deschişi uneori [...]. Parcă nu e niciun control asupra gândurilor mele, se iscă singure – pare atât de viu. Jumătate din timp eu nu sunt sigur dacă eu doar mam gândit sau dacă chiar s-a întâmplat[38].

Observaţi folosirea frecventă a lui „eu”. Whitty şi Lewin bănuiau că era posibil ca pacientul lor să sufere de crize complexe, dar în multe alte cazuri cu leziuni bilaterale ale girusului cingular anterior se raportează experienţe similare. Dau mai jos exemplul unei paciente de-ale mele[39]. E vorba de o femeie de 44 de ani care a suferit o hemoragie subarahnoidă. A

raportat următoarele:

> E ca şi cum gândirea mea devine reală – când mă gândesc la ceva, imediat văd lucrul acela întâmplându-se aievea în faţa ochilor, apoi devin şi foarte confuză şi uneori nu ştiu ce s-a întâmplat cu adevărat şi ce a fost doar un gând.

Ea a dat un exemplu:

> *Pacienta:* Stăteam în pat şi mă gândeam, apoi s-a întâmplat cumva pur şi simplu că soţul meu [decedat] era acolo şi-mi vorbea. Apoi m-am dus să le fac baie copiilor şi apoi, brusc, am deschis ochii şi „Unde mă aflu?” – şi sunt singură!
> *Eu:* Adormiseşi?
> *Pacienta:* Nu cred. E ca şi cum gândurile mele numai ce s-au transformat în realitate.

Conştiinţa este evident anormală în aceste cazuri, dar nu asta vreau să arăt. Nu contestă nimeni implicarea cortexului în procesarea conştientă. Ceea ce nu se poate demonstra e opinia că sinele conştient este

generat acolo.

Observăm comportamente energice, intenţionate, de răspuns şi emoţionale la oameni şi animale care nu mai au deloc cortex. Introspecţia sugerează că conştiinţa este intim implicată în aceste comportamente în cazul meu şi al tău (prezumtiv normale). Aşadar, în pofida problemei altor minţi, am avea nevoie de un motiv temeinic să credem că ea nu e implicată şi în aceste cazuri. Dar, când căutăm asemenea motive temeinice, găsim doar povara istoriei mediului academic, un model preneurologic al minţii care s-a întâmplat să fie folosit ca şablon de primii exploratori ai creierului. Astăzi, foarte puţin oameni de ştiintă ar susţine versiunea empirismului lui Hume sau ideile lui Meynert din volumul *Psychiatry.* Ba chiar opiniile lui Meynert sunt în mare parte luate în râs şi considerate o „mitologie a creierului”[40]. Şi totuşi, dogma conform căreia cortexul este organul minţii a devenit ipoteza fondatoare a unui întreg domeniu al medicinei.

Bineînţeles, cortexul este implicat în multe funcţii cognitive, inclusiv limbajul. Prin urmare, nu putem spera să primim rapoarte verbale introspective de la oameni la care este absent în totalitate. Dar avem mărturii nemijlocite de la pacienţi ale căror afecţiuni le permit să facă relatări verbale reflexive, chiar dacă le lipsesc părţile specifice ale cortexului despre care se presupune că dau naştere conştiinţei de sine generale. De nenumărate ori, aceşti pacienţi afirmă că sunt conştienţi şi îşi declară „existenţa” introspectivă. Oare mint? Dacă da, de ce? Şi oare ce-o fi însemnând afirmaţia că o persoană fără un sine minte cu privire la

propria ipseitate? Neuroştiinţa cognitivă e aproape incoerentă în acest punct – semn clar că a luat-o într-o direcţie greşită.

Eu consider că, în baza dovezilor existente, teoria conform căreia conştiinţa survine exclusiv în cortex nu se confirmă. Nu există motive temeinice să credem că cortexul dă naştere existenţei conştiente, în sensul în care o trăim şi eu, şi tu în mod obişnuit, dar sunt multe motive temeinice să concluzionăm că această afirmaţie nu e valabilă. Trebuie să căutăm în altă parte sursa fiinţei noastre.

*1 E o ironie că rezultatul unor astfel de experimente a contribuit în cele din urmă la schimbarea atitudinilor noastre faţă de etica unor asemenea studii.

*2 Noţiunea filosofică de zombi e diferită de cea hollywoodiană: aici e vorba de o creatură umanoidă imaginară ce acţionează din toate punctele de vedere *ca şi cum* ar fi conştientă şi, din acest motiv, este nediferenţiabilă de o persoană normală, dar de fapt îi lipseşte dimensiunea interioară a experienţei subiective.

Care este obiectul experienţei?

Trebuie să fii conştient de ceea ce percepi şi înveţi ca să percepi şi să înveţi? Poate că bunul-simţ spune da. Filosofii empirişti au spus da. Dar răspunsul este nu.

Ideile empiriştilor au condus la concepţia clasică din neuroanatomie conform căreia percepţia (şi engramele derivate din ea) este ingredientul de bază al conştiinţei. Dar dovezile ştiinţifice care arată că *nu suntem constienti de r r / / majoritatea lucrurilor pe care le percepem şi le învăţăm* sunt acum copleşitoare. Percepţia şi memoria nu sunt funcţii cerebrale inerent conştiente. În acest sens, bunul-simţ s-a înşelat. Se dovedeşte că tot ce face mintea (cu excepţia unui singur lucru, după cum vom vedea) poate fi făcut destul de bine şi în mod inconştient.

E ciudat că această idee îi poate fi atribuită lui Sigmund Freud.

Deşi toată viaţa a susţinut teoria conform căreia conştiinţa survine exclusiv în cortex, Freud a fost unul dintre primii specialişti în neuroştiinţe care au pus sub semnul întrebării distincţia făcută de Meynert,

profesorul lui, între părţile „psihice” şi cele „nonpsihice” ale creierului[1]. După ce a analizat datele disponibile, a concluzionat în 1891 că opiniile lui „Munk şi ale altor cercetători care pleacă de la ideile lui Meynert [...] nu mai pot fi susţinute”[2]. Mai exact, nu a găsit nicio deosebire anatomică clară între căile „voluntare” şi cele „reflexe” postulate de Meynert. De asemenea, a demonstrat (numărând fibrele nervoase implicate) că imaginile mnezice corticale erau generate numai după o serie de legături intermediare dintre ele şi periferia senzorială. Aceste legături se aflau în regiunile presupuse a fi *nonpsihice* ale creierului şi reduceau volumul de informaţii transmise în fiecare etapă. Freud a dedus că aceste legături subcorticale trebuie să aibă vreun efect asupra informaţiilor senzoriale pe care le prelucrează. De observat metafora lui uimitoare, concepută într-o eră predigitală:

> Ştim doar că fibrele care ajung la cortexul cerebral după trecerea prin ţesuturile cenuşii [cele subcorticale] încă păstrează o anumită relaţie cu periferia corpului, dar nu mai pot transmite o imagine care să-i semene din punct de vedere topologic. Ele conţin periferia corpului în acelaşi fel în care [...] un poem conţine alfabetul, într-o rearanjare totală, servind unor scopuri diferite, cu numeroase legături între elementele topologice individuale, prin care unele dintre ele pot fi redate de mai multe ori, iar altele deloc[3].

În cazul în care cortexul nu este conectat direct la

periferia corpului, ci, mai degrabă, prin legături subcorticale intermediare, imaginile mnezice depozitate în cortex nu pot fi proiecţii propriu-zise ale lumii exterioare. Ele sunt probabil produsele finale ale unei prelucrări a informaţiilor în mai multe etape. Întrucât această prelucrare culminează cu imaginile mnezice corticale, partea subcorticală a prelucrării trebuie să genereze, în vreun sens, *versiuni preliminare* ale imaginilor mnezice. Prin urmare, legăturile subcorticale constituie probabil o parte din procesul mintal pe care îl numim „apercepţie”. Freud afirma că nu are sens să facem o distincţie artificială între părţile subcorticale şi cele corticale ale prelucrării informaţiilor şi să pretindem că numai produsul final este „psihic”:

> Se justifică să luăm o fibră nervoasă, care pe toată lungimea traseului ei este doar o structură fiziologică supusă modificărilor fiziologice, să-i cufundăm capătul în psihic şi să furnizăm acestui capăt o idee sau o imagine mnezică?[4]

Întrucât Freud a presupus, ca toată lumea din perioada aceea, că numai procesele corticale sunt *conştiente,* acest lucru l-a determinat să susţină ideea că percepţia şi învăţarea trebuie să includă etape preliminare *inconştiente* care sunt la fel de „psihice” ca şi cele conştiente. Adică engramele inconştiente (subcorticale) trebuie să fie la fel de psihice ca şi cele corticale: şi ele fac parte din funcţia pe care o numim „memorie”, deşi le lipseşte conştiinţa. Freud a concluzionat într-o scrisoare către Wilhelm Fliess: „În

esenţă nouă în teoria mea este teza că memoria e prezentă nu o dată, ci de mai multe ori, că e afişată în diferite tipuri de indicii”[5].

A identificat cinci etape succesive în prelucrarea informaţiilor: „percepţia”, „urma perceptivă”, „etapa inconştientă”, „etapa preconştientă” şi „etapa conştientă” (vezi figura 8). Diferenţa crucială dintre etape consta nu în faptul că etapa inconştientă era corporală, iar cea preconştientă era psihică, ci că tipul preconştient de procesare a memoriei putea fi refăcut în conştiinţă, dar nu şi tipul inconştient[6]. Cu alte cuvinte, *doar o parte a psihicului este conştientă.*

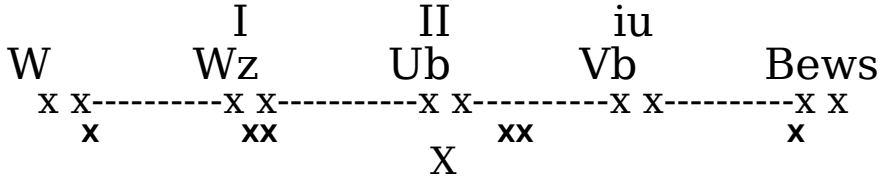


Figura 8. Prima diagramă a sistemelor de memorie realizată de Freud, ataşată scrisorii sale către Wilhelm Fliess *(W* = percepţie, *Wz* = urma perceptivă, *Ub* = etapa inconştientă, *Vb* = etapa preconştientă, *Bews* = etapa conştientă)

În 1895 Freud a făcut o observaţie similară despre căile nervoase care pleacă din *interiorul* corpului. Nici în acest caz, a afirmat el, nu are sens să spunem că informaţiile corporale viscerale devin „psihice” numai

când ajung la cortex. Întrucât aceste informaţii (care transmit senzaţii precum foamea sau setea) solicită psihicului să lucreze ca urmare a conexiunii cu corpul, trebuie să înglobăm şi aceste solicitări în imaginea noastră despre psihic. Chiar dacă nu îţi conştientizezi „pulsiunile” biologice, cum le numea Freud, cu siguranţă ele fac parte din psihicul tău. Aminteşte-ţi din capitolul 2 că Freud a sugerat chiar că solicitările corpului furnizează „motorul mecanismului psihic” şi că prozencefalul (al cărui cortex reprezintă lumea exterioară) este doar un „ganglion simpatic”. Cu alte cuvinte, el a concluzionat că atât imaginile mnezice conştiente, cât şi cele inconştiente sunt formate „în simpatie” cu solicitările din partea corpului – că ne reprezentăm şi învăţăm despre lumea exterioară numai prin intermediul percepţiei, deoarece acolo trebuie să ne satisfacem nevoile biologice.

La această perspectivă asupra creierului ştiinţa modernă adaugă multe detalii despre care Freud nu avea nicio idee şi îl corectăm în unele puncte importante. Chiar şi aşa, concluzia lui fundamentală este în general acceptată astăzi în neuroştiinţe: creierul îndeplineşte o mare diversitate de funcţii *psihice* care nu intră în conştiinţă. Titlul unei celebre treceri în revistă a literaturii relevante semnate de John Kihlstrom, un specialist modern în ştiinţele cognitive, e cât se poate de sugestiv: există într-adevăr „percepţie fără conştientizarea a ceea ce este perceput, învăţare fără conştientizarea a ceea ce este învăţat”[7]. O altă analiză bine-cunoscută, scrisă de John Bargh şi Tanya Chartrand, rezumă perspectiva pe care o avem astăzi

sub un titlu mai poetic, „Insuportabilul automatism al fiinţei”:

> Cea mai mare parte a vieţii psihice din fiecare clipă trebuie să apară prin mijloace nonconştiente, dacă e să apară [...]. Reglarea conştientă şi intenţionată a comportamentului, evaluărilor, deciziilor şi stărilor noastre emoţionale necesită un efort considerabil şi este relativ lentă. Mai mult, ea pare că necesită o resursă limitată care se epuizează rapid, aşa că actele de autoreglare conştientă nu pot avea loc decât în mică măsură şi pentru scurt timp. Pe de altă parte, [procesele psihologice...] nonconştiente sau automate sunt neintenţionate, facile, foarte rapide şi multe dintre ele se pot desfăşura în orice moment. Cel mai important, ele nu presupun efort, funcţionează încontinuu, ghidându-l pe individ în siguranţă pe parcursul zilei[8].

Totuşi, recunoaşterea modernă a ubicuităţii funcţionării psihice inconştiente nu s-a bazat pe teoriile lui Freud. La aceasta s-a ajuns într-un mod cu totul independent, pe baza altor descoperiri neurologice şi dovezi experimentale. Un punct de cotitură a fost cazul lui H.M., care a fost publicat în 1957. Ca urmare a rezecţiei chirurgicale a hipocampilor (regiunea cortexului responsabilă cu codificarea amintirilor declarative, cea care îi producea crize epileptice), era complet incapabil să- şi amintească vreun eveniment petrecut după data intervenţiei. Cu toate acestea, neuropsihologii care l-au examinat în mod repetat de atunci (1953) până la moartea sa, în 2008, au observat

că scorurile sale la testele psihometrice s-au îmbunătăţit semnificativ de-a lungul anilor, demonstrând efecte clare ale exerciţiului. A învăţat cum să se descurce la teste, deşi nu îşi amintea în mod conştient că le mai făcuse[9].

Observaţii asemănătoare a făcut anterior şi Édouard Claparède, deşi nu şi-a dat seama de importanţa lor. A raportat cazul unei femei amnezice care dădea mâna cu el în fiecare zi ca şi cum nu îl mai întâlnise niciodată (ca pacientul meu, domnul S.). Claparède s-a hotărât să-şi ascundă o piuneză în palmă, ca să vadă dacă ea va fi la fel de dornică să-l salute după ce a fost înţepată. După această experienţă, pacienta a refuzat să dea mâna cu el, deşi nu îşi amintea în mod conştient de evenimentul dureros. Când a întrebat-o de ce refuză, ea a dat doar explicaţii evazive, precum: „Oare doamnele nu au dreptul să nu dea mâna cu domnii?”[10].

O procesare inconştientă echivalentă a fost demonstrată şi pentru apercepţie. Înainte ca comisiile de control instituţional să-i evalueze etica protocoalelor de cercetare medicală, Roger Sperry a raportat cazul unei femei ale cărei emisfere cerebrale fuseseră separate chirurgical pentru a controla o epilepsie netratabilă. Când emisfera sa dreaptă izolată a fost expusă la o succesiune rapidă de imagini pornografice, pacienta a chicotit şi a roşit, deşi emisfera ei stângă (verbală) nu a putut „să declare” în mod conştient despre ce sentimente era vorba. Ea a putut să facă doar nişte afirmaţii tangenţiale pentru a-şi explica stânjeneala, precum „Ce maşinărie grozavă aveţi,

domnule Sperry!”[11].

Aici e important să observăm că procesele mintale inconştiente nu se limitează la structurile cerebrale subcorticale. În cazul lui Sperry, de exemplu, pacienta era incapabilă să conştientizeze informaţiile perceptive stânjenitoare primite, procesate şi recunoscute de cortexul ei cerebral drept.

Faptul că cortexul poate îndeplini asemenea funcţii fără conştientizare a fost demonstrat în mod concludent de multe studii experimentale. Într-unul din ele, participanţii la cercetare au fost expuşi la o succesiune foarte rapidă de cuvinte pozitive şi negative – atât de rapidă, încât nu au fost conştienţi că au văzut ceva. Comportamentul lor ulterior a fost în mod clar influenţat de cuvintele pe care pretindeau că nu le văzuseră: de exemplu, după expunerea la o succesiune rapidă de atribute negative asociate cu o fotografie a chipului A şi de atribute pozitive asociate cu chipul B, participanţii au manifestat o preferinţă pentru chipul B, deşi nu ştiau de ce îl preferă[12]. Asta arată că cuvintele pozitive şi cele negative trebuie să fi fost văzute, citite *şi înţelese* în mod inconştient. Întrucât citirea implicând comprehensiunea este o funcţie exclusiv corticală – exact de tipul celor considerate tipic „psihice” de anatomiştii clasici –, nu putem concluziona decât că *funcţiile corticale nu sunt inerent conştiente.*

Aceasta este opinia general acceptată în ştiinţa cognitivă de astăzi: psihicul interpretat drept imagini mnezice (acum numite „reprezentări”) nu este conştient în mod intrinsec. Partea din creier care generează „conţinutul” conştiinţei („conştiinţa ca experienţă”) –

cortexul – poate să facă exact acelaşi lucru şi în lipsa experienţei conştiente. Dar, dacă aşa stau lucrurile, atunci ce anume face ca funcţiile sale inconştiente să fie *psihice?* Prin ce se deosebeşte cortexul – aşa-zisul organ al minţii – de alte dispozitive de procesare a informaţiilor precum telefoanele mobile? Din nou, riscăm să devenim incoerenţi.

Revenim astfel la marea întrebare cu care am încheiat capitolul 3: dacă nu din cortex vine conştiinţa noastră, atunci de unde? Şi, mai mult, dacă cea mai mare parte a „vieţii psihice din fiecare clipă” se desfăşoară fără experienţă conştientă, atunci de ce *uneori* presupune experienţă conştientă? De ce toată această prelucrare a informaţiilor nu se desfăşoară nonconştient?

Istoria neuroştiinţei cognitive este un cimitir al teoriilor care au încercat să specifice funcţia conştiinţei. Ele au în comun ipoteza că funcţia sa este de „a lega” multitudinea fluxurilor de informaţie ce se întind prin tot creierul în întregul coerent care ne caracterizează experienţa conştientă. De exemplu, cititul, recunoaşterea facială, percepţia culorilor, percepţia mişcării, recunoaşterea obiectelor, percepţia spaţiului şi altele asemenea, toate acestea se întâmplă simultan în arii localizate în diverse părţi ale creierului. Cum anume se îmbină ele în imaginile vizuale unificate pe care le percepem în mod obişnuit – cu culoarea şi mişcarea unui chip, de exemplu, întâmplându-se exact în locurile potrivite?

Meynert a afirmat că această funcţie de conexare era

îndeplinită de fibrele transcorticale, care „asociază” între ele imagini mnezice. Au trecut mai bine de o sută de ani, dar noi cam tot la această ipoteză am rămas. James Newman şi Bernard Baars au sugerat că un „spaţiu de lucru global” unificat este generat în cortex de talamus, făcând fragmentele disparate de informaţie accesibile în mod global experienţei. Stanislas Dehaene şi Lionel Naccache au adăugat că ariile corticale de asociere prefrontale şi parietale integrează activităţile zonelor senzoriale primare în acest spaţiu de lucru. Gerald Edelman a introdus „circuitele de reintrare” talamocorticale drept funcţia esenţială prin care informaţia integrată este trimisă înapoi la nivelurile anterioare ale procesării perceptive. Giulio Tononi a evidenţiat procesul rezultant de prelucrare a informaţiilor „integrate masiv”, afirmând că tocmai gradul de integrare a acestor fragmente este componenta-cheie; conştiinţa este o funcţie a *cât de multe* informaţii se integrează. Francis Crick şi Christof Koch au emis ipoteza că sincronia oscilaţiilor gama din cortex conexează şi stochează experienţele – cu alte cuvinte, că integrarea s-ar putea petrece în timp, şi nu în spaţiu. Rodolfo Llinás a sugerat în mod similar că activitatea talamocorticală se sincronizează sub 40 Hz. Şi aşa mai departe[13].

Niciuna dintre aceste teorii nu ne spune de ce sau în ce mod conexarea informaţiilor – prin asociere, oscilare, sincronizare, reintrare, integrare masivă şi altele asemenea – dă naştere *cu necesitate* experienţei. De ce, mai precis, trebuie resimţit în mod conştient conţinutul „spaţiului de lucru global” pentru prelucrarea

informaţiilor? Oare conexarea, stocarea, sincronizarea, integrarea masivă şi aşa mai departe nu apar şi în cadrul prelucrării *inconştiente* a informaţiilor? Calculatoarele generează spaţii de lucru globale şi integrează masiv informaţii tot timpul, atunci când sunt interconectate prin internet. Şi atunci de ce n-ar fi internetul conştient?

Un aspect îngrijorător, ce decurge direct din aceste teorii contemporane, este că tot mai mulţi specialişti respectabili în neuroştiinţe (precum Koch şi Tononi) sugerează că acest lucru *ar fi* posibil. Asta din cauză că teoriile lor îi obligă să accepte această posibilitate ciudată. Procedând astfel, ei urmează curentul de gândire „panpsihist” iniţiat de Thomas Nagel, conform căruia s-ar putea ca *toate* lucrurile să fie (puţin) conştiente[14].

Ar trebui să evaluăm aceste teorii în lumina observaţiilor despre care am vorbit mai sus, în sensul că conştiinţa persistă în lipsa cortexului. Newman şi Baars îşi localizează „spaţiul de lucru global” în cortexul cerebral şi totuşi animalele supuse intervenţiei de decorticare cerebrală şi copiii născuţi fără cortex par să fie conştienţi. Cel puţin, dovezile că aceste animale şi aceşti copii sunt fiinţe conştiente sunt mult mai convingătoare decât dovezile că internetul ar fi conştient. Mai mult, adulţii care pot să vorbească şi al căror substrat aşa-zis esenţial pentru prelucrarea informaţiilor integrate masiv a fost distrus de boală ne spun că simţul lor al sinelui persistă în lipsa substratului respectiv. Această afirmaţie este valabilă în cazul pacientului meu W., căruia îi lipseau complet lobii

prefrontali.

Nimic din toate acestea nu ar trebui să ne surprindă dacă vom combina asemenea observaţii clinice cu dovezile experimentale analizate anterior, în sensul că cortexul efectuează *inconştient* cea mai mare parte a prelucrării informaţiilor (cum sunt cititul şi recunoaşterea facială).

În dovezile pe care le-am prezentat există ceva remarcabil la care aş vrea să mă refer acum. Când pacienta lui Claparède a refuzat să dea mâna cu el, ea nu ştia de ce. Totuşi, probabil *a simţit* o anumită aversiune faţă de acest gest. Probabil avea vreo bază subiectivă când a refuzat să dea mâna cu el, deşi nu avea acces la cauza obiectivă (amintirea piunezei). Acelaşi lucru e valabil şi în cazul pacientei lui Sperry: s-a simţit stânjenită, dar nu ştia de ce. Cu alte cuvinte, nu era conştientă de cauza obiectivă a sentimentului ei (imaginile pornografice), dar a conştientizat sentimentele subiective care le-au însoţit (stânjeneala).

Acelaşi lucru s-a întâmplat probabil şi în experimentul cu succesiunea rapidă de cuvinte: participanţii au simţit probabil o preferinţă pentru chipul B, deşi le lipsea conştientizarea explicaţiei acestui fapt. Acelaşi lucru pare să fie valabil şi în cazul fetiţei cu hidroanencefalie din figura 5: ea a manifestat o plăcere subiectivă când frăţiorul său i-a fost aşezat în poală, deşi nu avea cum să ştie cauza obiectivă a emoţiei sale, fiind complet lipsită de imagini corticale.

Prin urmare, în toate aceste cazuri, când am spus că pacienţii şi participanţii la cercetare pe care i-am descris nu erau „conştienţi” de cauzele acţiunilor lor, m-

am exprimat prea vag. Ei nu erau conştienţi de anumite percepţii sau amintiri – reprezentări –, dar *sentimentele* lor persistau. Exista încă „ceva cum e” să fii ei, care le determina judecăţile de valoare. Erau conştienţi de sentimentele lor, numai că nu le conştientizau sursa.

Acest lucru ne oferă un indiciu important cu privire la ce înseamnă în mod fundamental capacitatea de a simţi. Aparent singurele printre funcţiile psihice, sentimentele sunt *cu necesitate* conştiente. Cine a auzit vreodată de un sentiment ce nu are o calitate subiectivă? Care ar fi rostul unui sentiment dacă nu l-ai simţi? Chiar Freud a acceptat că sentimentele *trebuie să fie* conştiente:

> Din esenţa unui sentiment face totuşi parte faptul că el e resimţit, că e cunoscut, aşadar, de către conştiinţă. Posibilitatea unui inconştient nu s-ar pune absolut deloc atunci când vorbim de sentimente, senzaţii şi afecte*[1] 15.

Îmi dau seama că unele persoane (inclusiv psihanalişti) resping această afirmaţie cu argumentul că există emoţii inconştiente. Voi reveni curând la această chestiune. Deocamdată, aş vrea să fie cât se poate de clar cu ce sens folosesc termenul „sentiment”: mă refer la acel aspect al unei emoţii (sau al oricărui afect) pe care îl *simţi.* Mă refer la ceea ce simţi ca atare. Dacă nu simţi ceva, atunci nu vorbim de un sentiment.

Cu câteva paragrafe mai sus am spus că pacienta lui Claparède trebuie să fi simţit o oarecare aversiune, pacienta lui Sperry trebuie să fi simţit stânjeneală,

participanţii la testul cu succesiunea rapidă de cuvinte trebuie să fi simţit o preferinţă şi aşa mai departe. M-am exprimat astfel pentru că specialiştii în neuroştiinţe de obicei nu pun întrebări despre experienţe subiective. Acesta este singurul motiv pentru care nu ştim ce au simţit aceşti subiecţi de cercetare. Însă eu (ca şi Freud şi Sacks) iau foarte în serios relatările introspective ale pacienţilor. În acest fel, sper să evit erori precum confundarea visului cu somnul REM.

Dacă o sută de oameni dintr-o sută raportează o senzaţie de durere când îi ciupeşti de mână, nu este iraţional să concluzionezi că ciupirea mâinii cauzează durere, chiar dacă trebuie să te bazezi pe relatări introspective în acest sens în fiecare situaţie. Această afirmaţie este chiar mai valabilă pentru observaţiile pe care le poţi replica în cazul tău, când ai acces direct la fenomenul în cauză. Dacă mă doare când mă ciupesc de mână, atunci am o experienţă personală a ceea ce vor să spună alţii când raportează „durere”.

Sentimentele sunt dificil de studiat din cauză că sunt inerent subiective, dar (cu voia behavioriştilor) asta nu ne îndreptăţeşte să le ignorăm. Dacă excludem sentimentele din descrierea creierului, nu vom înţelege niciodată cum funcţionează acesta. Rolul jucat de sentimente în confabulaţiile domnului S. ilustrează foarte bine această idee.

Când ne uităm la ce apare în conştiinţa noastră, observăm câteva categorii generale de conţinut. Găsim, desigur, „reprezentările” lumii exterioare – atât

percepţii, amintiri ale ei, cât şi gânduri despre ea. Filosofii acordă multă atenţie acestor reprezentări, iar modelul empirist al minţii a fost conceput pentru a le explica. Dar în conştiinţă mai găsim şi altceva.

Găsim şi sentimente – sentimente despre ce se întâmplă în lume, sentimente despre gândurile noastre în privinţa lumii, sentimente (mai ales) despre noi înşine, inclusiv sentimente ce par să fie relatări despre starea propriului corp. Există şi sentimente vagi, sentimentele şi dispoziţiile afective care descriu modul în care resimţim lumea şi care ne modelează comportamentul în interiorul ei. Uneori, ele trec drept senzaţii corporale, însă multe dispoziţii afective par a nu se datora nici stării corpului, nici vreunui alt lucru pe care îl putem identifica în lumea exterioară. Oare nu e conştiinţa plină de asemenea sentimente? Şi totuşi, e uimitor cât de frecvent sunt trecute cu vederea de specialiştii în neuroştiinţe care caută o explicaţie a conştiinţei[16].

E uluitor că filosofii empirişti şi moştenitorii lor din domeniul ştiinţei, behavioriştii şi cognitiviştii, au ignorat sentimentele[17]. Behavioriştii au afirmat că întreaga învăţare este guvernată de „recompense” şi „pedepse”, dar nu ne-au spus niciodată ce sunt acestea. Au realizat experimente riguroase care au generat „legea efectului”. Conform acesteia, dacă un comportament este urmat constant de recompense, se va intensifica, iar dacă e urmat constant de pedepse, se va diminua. Acest proces al învăţării din experienţă a fost numit „condiţionare”.

Edward Thorndike, autorul legii efectului, a vrut să

demonstreze că animalele învaţă prin încercare şi eroare, nu judecând. Dar, de fapt, legea lui Thorndike echivalează cu o lege a *afectului*[18], deoarece sugerează că repetăm (şi noi, şi animalele) comportamentele care ne fac să ne simţim bine şi că le evităm pe cele care ne fac să ne simţim rău. Aşadar, legea efectului nu e în esenţa sa decât „principiul plăcerii" despre care vorbea Freud. Dar behavioriştii nu puteau să accepte existenţa a ceva atât de subiectiv precum sentimentele. E bine-cunoscut că B.F. Skinner, de exemplu, a declarat: „«Emoţiile» sunt exemple excelente ale cauzelor fictive pe seama cărora punem în mod obişnuit comportamentul"[19].

Felul cum şi-a formulat iniţial Thorndike legea dă în vileag minciuna: „Este mai probabil ca răspunsurile care produc un efect satisfăcător într-o anumită situaţie să apară din nou în situaţia respectivă şi mai puţin probabil ca răspunsurile care produc un efect neplăcut să apară din nou în situaţia respectivă"[20]. Cuvintele „satisfăcător" şi „neplăcut" au fost mai apoi eliminate şi înlocuite cu „de întărire" şi „punitiv". Iată de ce:

> În psihologie, termenii noi, „de întărire" şi „punitiv", sunt folosiţi altfel decât în sens colocvial. Existenţa a ceva ce întăreşte un comportament creşte probabilitatea ca acel comportament să apară din nou, iar existenţa a ceva ce pedepseşte un comportament reduce probabilitatea ca acel comportament să apară din nou[21].

Această definiţie a termenilor e cu totul lipsită de

conţinut, după cum şi trebuie să fie, întrucât cuvântul esenţial este „ceva”. Aceeaşi afirmaţie e valabilă în cazul termenilor behaviorişti „întărire pozitivă” şi „întărire negativă”: ce anume face ca o întărire să fie pozitivă sau negativă dacă nu un sentiment?

Probabil se intenţiona ca semnificaţia „recompensei” şi „pedepsei” să fie că valoarea este intrinsecă *stimulului* – şi nu receptorului stimulului. Dacă un cal se apropie de mine şi eu îi dau un cub de zahăr, va exista (conform legii efectului) o probabilitate mai mare să se apropie de mine din nou, dar, dacă îl stropesc cu suc de lămâie în ochi, această probabilitate scade. Conform lui Thorndike, cubul de zahăr şi lămâia devin *ele însele* în acest fel recompense sau pedepse ale comportamentului calului; nu e nevoie să luăm în calcul senzaţiile pe care le evocă, dacă există aşa ceva. Bineînţeles, acesta este un raţionament greşit, care plasează eronat forţa cauzală, având drept consecinţă problema de care s-a plâns Sacks: un psihic lăsat fără intenţionalitate. Una e să tratezi creierul ca pe o cutie neagră din raţiuni metodologice şi cu totul altceva să atribui lucrurilor din afara cutiei puteri cauzale pe care de fapt ele nu le au, iar apoi să concluzionezi că în cutie nu se întâmplă nimic.

Ideea că sentimentele sunt „fictive” are multe efecte nocive asupra ştiinţei. De exemplu, în cea mai mare parte a secolului trecut, atunci când erau studiate mecanismele fiziologice de bază ale echilibrului energetic, behavioriştii au interzis folosirea cuvintelor „foame” şi „saţietate”, întrucât asemenea lucruri nu pot fi văzute sau atinse. Behavioriştii vorbeau numai despre

„stimulente” şi „recompense” asociate cu comportamentul alimentar. Aceasta nu e doar o chestiune de semantică. Dacă un cuvânt precum „foame” nu poate fi folosit, atunci cum putem să înţelegem rolul pe care îl joacă ea în reglarea mâncatului? Oare acest lucru nu ar putea să întârzie, la rândul său, dezvoltarea unor tratamente pentru reducerea senzaţiei de foame în cazurile de obezitate? De fapt, comportamentul alimentar este reglat de două mecanisme cerebrale ce interacţionează între ele: un sistem „homeostatic”, ce reglează rezervele de energie, şi un sistem „hedonist”, ce acţionează asupra *apetitului*[22]. Şi, la fel ca în cazul unor afecte corporale precum foamea, oare nu ar fi posibil ca interzicerea unor cuvinte denumind emoţii precum „tristeţe” şi „frică” să întârzie dezvoltarea unor tratamente antidepresive şi anxiolitice?— Dacă sentimentele chiar există, atunci cu siguranţă că au corelative fiziologice care ne vor scăpa dacă nu le luăm în considerare.

În cele ce urmează mă voi despărţi de behaviorişti. Urmându-l pe Panksepp, primul specialist în neuroştiinţe care a avut curajul să folosească cuvinte precum „foame” ca să explice reglarea echilibrului energetic atât la oameni, cât şi la alte animale, voi accepta că sentimentele există *cu adevărat*[24]. Voi presupune că ştii ce înseamnă să simţi că îţi este sete sau că eşti trist, somnoros, amuzat, încrezător sau nesigur. Această ipoteză nu se justifică mai puţin decât alte inferenţe ştiinţifice despre lucruri din natură. Şi poate fi testată prin metodele obişnuite, prin predicţii falsificabile.

Sentimentele sunt reale şi ştim de existenţa lor pentru că sunt peste tot în conştiinţa noastră. De fapt, ele sunt, din motivele pe care le voi explica în continuare, izvorul fiinţei conştiente – într-un sens care nu mi se pare deloc metaforic. De la originea lor în unele dintre cele mai străvechi straturi ale creierului, ele irigă solul sterp al reprezentărilor inconştiente şi le aduc la viaţa psihică.

*1 Sigmund Freud, „Inconştientul”, traducere de G. Lepădatu, în *Opere esenţiale,* vol III: *Psihologia inconştientului.* Bucureşti: Trei, e-book.

5

Sentimentele

Spre încântarea medicilor lui şi a noastră a tuturor, fratele meu Lee şi-a revenit cu bine după leziunea la creier şi s-a adaptat excelent. Deşi putea fi indus în eroare cu uşurinţă de oameni care căutau să profite de el, din punct de vedere fizic era masiv şi puternic şi s-a folosit de acest lucru cu succes. Nu a mai fost niciodată exact acelaşi ca înainte de accident, dar şi-a continuat viaţa.

Mie mi-a fost mai greu să mă adaptez.

Când aveam cinci ani, tata i-a cumpărat lui Lee un ceas de mână, pe care i-a promis că i-l dă când va învăţa să spună cât e ora. A urmat o lecţie scurtă. Lee s-a chinuit cu întrebări despre ce înseamnă când acul mic e la 9 şi cel mare la 11. Observându-l din capătul opus al încăperii, am dat eu rapid răspunsurile. Tata mi-a spus să tac. În seara aceea i-a dat ceasul lui Lee, deşi fratele meu încă nu se descurca bine. Şi ceasul meu? mă gândeam.

Pe măsură ce am mai crescut, asemenea sentimente copilăreşti au făcut loc vinovăţiei. Eu mă descurcam

bine la şcoală. Lee a fost norocos că a terminat-o. Pur şi simplu, pentru mine lucrurile păreau să fie mai uşoare. Îmi doream ca familia mea să poată vorbi despre această problemă, ca să găsească strategii care să mă ajute să nu îi creez lui neplăceri. Totuşi, mama nu suporta să se discute despre accident, iar tata nu era genul de persoană cu care să pot discuta *orice* sincer – şi în niciun caz despre cum îşi făcea el treaba de cap al familiei.

Astăzi pare ceva evident, dar, când m-am hotărât să studiez neuropsihologia, nu îmi trecea prin cap că ar putea să aibă vreo legătură cu accidentul lui Lee. Îmi amintesc însă că în copilărie mă gândeam că singurul lucru pe care merită să-l fac în viaţă este să aflu ce înseamnă „să exişti”. Abia în perioada în care am făcut studii de psihanaliză, după ani de zile, când deja lucram ca neuropsiholog, am reuşit să pun totul cap la cap. În mod clar, cariera pentru care optasem era un compromis între ambiţie, pe de o parte, şi vinovăţie, pe de altă parte. Dacă deveneam neuropsiholog, puteam să am succes din punct de vedere academic şi, în acelaşi timp, să ajut oameni aflaţi în dificultate. De asemenea, probabil că asta explică de ce mă simţeam atât de frustrat de neuropsihologia pur academică şi de ce m-am hotărât să devin clinician şi, ocazional, chiar terapeut pentru neurorecuperare – roluri profesionale pe care specialiştii în neuroştiinţe tind să le desconsidere.

Întrerup puţin şirul amintirilor. Ţi se par toate acestea o explicaţie plauzibilă a deciziei mele? Aceasta nu este o întrebare-capcană: cred că amintirile ascunse

ale acestor experienţe explică motivele alegerilor mele în carieră. Totuşi, în general, oare nu aşa par adesea că ne motivează sentimentele noastre, undeva sub pragul a ceea ce conştientizăm? E vorba de un loc comun al explicaţiilor psihologice – un element de freudism atât de bine integrat în bunul-simţ, încât puţini s-ar gândi să-l conteste. Păi, astfel, e ciudat să afirmi că sentimentele sunt conştiente prin definiţie: că sunt, într-un anumit sens, esenţa conştiinţei. Cum să fie posibil aşa ceva?

Într-o zi din 1985, după ce mi-am încheiat studiile universitare, m-am pornit spre spitalul Baragwanath pentru prima mea zi ca neuropsiholog stagiar. Contextul psihosomatic e că eram îngrijorat: profesorul meu – Michael Saling – îşi luase anul sabatic şi plecase în Australia, eu rămânând fără supravegherea obişnuită. În plus, mă stresa şi faptul că m-am rătăcit în drum spre spital, în Soweto, un complex urban foarte periculos pentru băieţi albi ca mine.

Când, în cele din urmă, am găsit spitalul, uşurat că ajunsesem nevătămat, m-am repezit spre Secţia 7 (neurochirurgie) şi am întrebat de domnul Percy Miller, neurochirurgul consultant la care mi se spusese să mă prezint. Infirmiera-şefă m-a condus prin secţie. Mirosul era puternic: un amestec de fluide corporale, antiseptic şi ceva ce aducea a varză fiartă.

Miller semăna cu o pasăre. În timp ce stătea aplecat deasupra unui pacient căruia îi drena lichid cefalorahidian printr-un orificiu din cap, a început să-mi vorbească. „Eşti Mark Solms? Încântat de cunoştinţă. Mike ne-a spus că-i ţii locul. Mă bucur c-ai venit. Aşteaptă puţin. Termin imediat şi îţi fac turul secţiei.”

Am încercat să par calm şi profesionist. Să-i ţin locul? Neurochirurgul şi-a spălat mâinile, apoi m-a dus la primul pat. Până să-mi dau seama ce se întâmplă, a început să turuie istoricul clinic al pacientului. „Acesta este domnul cutare, 43 de ani, astrocitom temporal stâng, grad III. Îl operăm miercuri. Evaluează-i în linii mari funcţiile limbajului şi zi-ne şi nouă. Şi memoria, presupun; dumneata te pricepi mai bine decât mine.”

M-am gândit dacă să întreb ce e un astrocitom grad III.

Am trecut la patul următor. „Acesta este domnul cutare, 58 de ani. Are un macroadenom hipofizar. Îl operăm mâine, aşa că te rog să-l evaluezi mai întâi pe el. Facem abordare transsfenoidală. Mike consideră că aşa rezultatele cognitive sunt mai bune – ceea ce are sens.”

Şi tot aşa, de la un pat la altul. Când am ajuns la al cincilea sau al şaselea, era prea târziu să mărturisesc că nu înţelegeam mai nimic din ce mi se spunea. De asemenea, nu îmi puteam aminti detaliile cele mai importante ale primelor câteva cazuri. Abia dacă îmi aminteam câte ceva. Mi-am spus că voi reveni mai târziu ca să citesc fişele.

Am trecut prin nişte uşi batante şi am intrat în salonul pentru femei. Chirurgul vorbea în continuare: „Aceasta este doamna cutare, 36 de ani, cisticercoză”. Nici asta nu ştiam ce înseamnă. Următoarea pacientă era complet goală. Şi unii dintre bărbaţi erau goi, dar asta era altceva. M-am gândit că nu ar fi trebuit să văd aşa ceva. Oare ei nu ştiu că nu sunt medic? Pe parcursul

turului, observam expresiile faciale ale pacienţilor. Unul se uita la mine cu tristeţe, altul absent, unii neliniştiţi, mulţi nu mă priveau deloc.

Eram complet depăşit de situaţie. Oamenii ăştia presupuneau că ştiu ce fac. Şi, pentru că nu am comentat această presupunere, era clar că urma să fiu responsabil pentru moartea cuiva.

În salonul de pediatrie a fost şi mai rău. Primul copil era un caz grav de hidrocefalie: capul lui era cât al lui ET, de două ori mărimea normală. Mi s-a făcut greaţă, îmi era cald şi transpiram abundent. Al doilea copil avea un balon moale de carne care îi ieşea pe la ceafă. Se uita ţintă la perete. Percy Miller a spus ceva despre un „mielomeningocel”. Intensitatea vocii lui s-a atenuat, înlocuită de bătăile lente ale inimii mele. Mi s-a îngustat câmpul vizual şi, în timp ce începeam să nu îi mai văd gura, care vorbea, am avut un gând: o să mă lovesc cu capul de podea, o să-mi sparg capul şi-o să ajung într-unul din paturile astea. M-am simţit ciudat de uşurat, apoi mi-am pierdut cunoştinţa.

Acest exemplu al primei mele zile la serviciu ilustrează multe dintre caracteristicile esenţiale ale „afectului”, inclusiv faptul că, deşi nu ştim întotdeauna de ce simţim ce simţim, cu siguranţă ştim *ce* simţim. În acest caz, am simţit că urma o catastrofă, pe care am legat-o din punct de vedere intelectual de incompetenţa mea profesională, în spatele căreia se aflau tot felul de sentimente complicate referitoare la fratele meu, incluzând atât nevoia mea de a-l ajuta, cât şi o identificare cu el la un nivel profund. Dar cauza

imediată a leşinului meu a fost un reflex *corporal* primitiv, declanşat direct de emoţiile mele.

Sincopa vasovagală te face să leşini deoarece creierul reacţionează la ceva alarmant, de obicei la vederea sângelui sau la alt risc perceput de accidentare fizică. Acest declanşator (înregistrat de amigdală) activează nucleul solitar din trunchiul cerebral, ceea ce duce la scăderea bruscă a pulsului şi a tensiunii arteriale. Iar asta duce la un aport sangvin redus în creier; şi îşi pierzi cunoştinţa.

De ce avem acest reflex înnăscut? Răspunsul este: el reduce fluxul sangvin şi, astfel, opreşte o hemoragie, în anticiparea unei leziuni[1]. Acest reflex duce la leşin numai la oameni, din cauza staţiunii noastre verticale şi a creierului mare, ce necesită un efort cardiac mai mare. (Prin urmare, e puţin probabilă teoria că leşinul este o formă de „a face pe mortul”.) În linii mari, reflexul vasovagal creşte şansele să supravieţuieşti unei accidentări, nu le reduce; şi, în orice caz, fluxul sangvin cerebral se restabileşte imediat ce cazi la pământ.

Să luăm alt exemplu[2]. Controlul respirator este în mod normal automat: cât timp nivelurile de oxigen şi dioxid de carbon din sânge rămân în limite viabile, nu e nevoie să conştientizezi respiraţia ca să respiri. Totuşi, atunci când gazele sangvine depăşesc aceste limite normale, controlul respirator intră peste conştiinţă sub forma unei senzaţii acute numite „foame de aer”. Valorile neprevăzute ale gazelor sangvine sunt un semnal că *trebuie luate măsuri.* E o nevoie urgentă să eliberezi o obstrucţie a căilor respiratorii sau să ieşi dintr-o încăpere plină de dioxid de carbon. În acest

moment, controlul respirator este conştientizat printr-un sistem intern de avertizare pe care îl resimţim ca pe o *alarmă* – în mod specific, în acest caz, alarmă de sufocare.

Unele dintre cele mai simple forme de senzaţii – foamea, setea, somnolenţa, oboseala musculară, greaţa, frigul, urgenţa urinară, nevoia de a defeca şi altele asemenea – poate că nu par afecte, dar ele asta sunt. Stările afective se deosebesc de alte stări psihice prin faptul că au *valenţe* hedonice: se simt ca „bune” sau „rele”. În asta constă diferenţa dintre senzaţiile afective precum foamea şi setea şi cele senzoriale precum văzul şi auzul. Imaginea şi sunetul nu au o valoare intrinsecă, spre deosebire de senzaţiile afective[3].

„Binele” sau „răul” unei senzaţii îţi spune ceva despre starea nevoii biologice din spatele ei. Astfel, setea e ceva rău, iar potolirea ei e ceva bun, întrucât e necesar să-ţi menţii nivelul de hidratare în limitele viabile pentru supravieţuire. Aceeaşi afirmaţie e valabilă în cazul senzaţiei neplăcute de foame în relaţie cu uşurarea plăcută pe care ţi-o dă mâncatul. Pe scurt, plăcerea şi neplăcerea îţi spun *cum te descurci* în legătură cu nevoile tale biologice. Valenţa reflectă sistemul de valori ce susţine toate aspectele vieţii biologice, şi anume că e „bine” să supravieţuieşti şi să te reproduci şi „rău” să nu faci asta.

Desigur, individul nu e motivat direct de aceste valori biologice, ci de senzaţiile subiective generate de ele – chiar dacă nu avem nicio idee care sunt valorile biologice de bază şi chiar dacă nu le susţinem din punct de vedere intelectual. De exemplu, mâncăm alimente

dulci pentru că au gust bun, şi nu pentru că tind să aibă un conţinut caloric ridicat, care este motivul biologic pentru care au gust bun. Afectele spun poveşti lungi despre evoluţie pe care nu le conştientizăm deloc. Ca în cazul reflexului vasovagal, conştientizăm numai senzaţiile.

Motivul pentru care folosesc cuvântul „neplăcere”, şi nu „durere” este că în creier există multe tipuri diferite de plăcere şi de neplăcere. Foamea e o senzaţie neplăcută şi te simţi bine când o atenuezi mâncând; intestinele balonate îţi dau o senzaţie neplăcută şi te simţi bine când o atenuezi defecând; durerea e o senzaţie neplăcută şi te simţi bine dacă te îndepărtezi de sursa ei. Acestea sunt afecte corporale, dar acelaşi lucru se aplică şi celor emoţionale. Detresa cauzată de separare e neplăcută şi răspundem prin a căuta reîntâlnirea. Frica e o senzaţie neplăcută şi scăpăm de ea îndepărtându-ne de pericol (şi, uneori, leşinând). Alarma de sufocare, foamea, somnolenţa şi frica sunt senzaţii neplăcute, dar din puncte de vedere diferite. Dar, dacă scăpăm de ele, ne simţim bine, tot din puncte de vedere diferite.

Senzaţiile diferite semnalează situaţii diferite semnificative din punct de vedere biologic şi fiecare dintre ele ne obligă *să facem* ceva diferit: urinarea nu poate satisface foamea, iar mâncatul nu poate uşura o vezică urinară plină. Să ne amintim ce a spus Damasio despre pacientul B.: „Raporta şi senzaţii de foame, de sete şi dorinţa de a excreta *şi avea comportamentele aferente”.* Ar fi fatal dacă am proceda altfel.

Senzaţiile ne determină să facem ceva *necesar.* În

acest sens, ele sunt indicatori ai cererilor de travaliu. În cazul foamei de aer, travaliul necesar trebuie să-ţi reechilibreze gazele sangvine. În cazul hipotermiei, travaliul trebuie să te readucă într-un interval de temperatură viabil. În cazul detresei cauzate de separare, travaliul trebuie să te ducă la reîntâlnirea cu cineva care îţi poartă de grijă. Şi aşa mai departe. În jargonul teoriei controlului, dezechilibrele gazelor sangvine, valorile suboptime ale temperaturii, lipsa celor care îţi poartă de grijă şi prădătorii ce se apropie sunt „semnale de eroare”, iar acţiunile pe care le generează sunt menite *să corecteze* erorile. Rezolvarea unui afect prin ceva de tipul satisfacerii înseamnă că o eroare a fost corectată cu succes, după care dispare de pe radarul conştiinţei.

Am ajuns din nou pe un teritoriu ce li se va părea ciudat de familiar psihanaliştilor. Ne amintim că Freud a definit „pulsiunea” ca „o măsură a cererii de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce este corporal”. Acum se vede cât de strâns legate sunt afectele de pulsiuni – ele sunt manifestările lor subiective. Afectele sunt modul în care devenim conştienţi de pulsiunile noastre, ele ne spun cât de bine sau cât de rău merg lucrurile în legătură cu nevoile specifice pe care le evaluează.

Iată pentru ce sunt afectele: ele transmit *ce* aspecte biologice merg bine sau rău pentru noi şi ne impulsionează să luăm măsuri. În acest sens, senzaţiile afective sunt diferite de cele perceptive. Filosofii sunt adesea neliniştiţi de posibilitatea a ceea ce ei numesc „qualia inversate”. Cum pot să ştiu că roşul pe care îl

văd eu este la fel ca roşul pe care îl vezi şi tu? Oare nu e posibil ca roşul meu să ţi se pară albastru? Problema altor minţi sugerează că nu putem şti niciodată, pentru că şi tu, şi eu am indica aceleaşi obiecte din lume şi le-am numi „roşii”. Dar ce e valabil despre percepţiile vizuale nu e valabil despre experienţa afectivă. Calitatea de a fi roşu nu *cauzează* nimic diferit de calitatea de a fi albastru, astfel încât le poţi schimba între ele în mod arbitrar fără consecinţe fizice.

Acest lucru nu e valabil în cazul unor senzaţii precum foamea în relaţie cu urgenţa urinară sau precum frica în relaţie cu detresa cauzată de separare. O senzaţie (frica) te constrânge să scapi de ceva; alta (detresa de separare) te constrânge să cauţi pe cineva. Senzaţia nu poate fi separată de starea corporală pe care o aduce cu sine[4]. Dacă le-ai schimba între ele, ai simţi o dorinţă irezistibilă de a scăpa de cineva care îţi poartă de grijă, dar lipseşte şi ai căuta plângând un prădător care te tot urmăreşte. Dacă ai schimba între ele calitatea subiectivă de a fi roşu şi cea de a fi albastru, nu ar exista consecinţe, dar schimbarea senzaţiei de frică cu detresa de separare (sau foamea cu urgenţa urinară) te-ar ucide[5].

Al doilea lucru de remarcat despre senzaţii e că *sunt întotdeauna conştiente.* O senzaţie de care nu eşti conştient nu este o senzaţie. Acest lucru e adevărat prin definiţie, după cum am spus mai sus, dar şi datorită unei trăsături specifice a fiziologiei creierului. Vom afla de ce stau lucrurile aşa în capitolul următor, unde discutăm despre

mecanismele cerebrale de „activare”. Deocamdată, vreau să te conving că senzaţiile sunt întotdeauna conştiente, fără excepţie. Nu vreau să spun că toate mecanismele de reglare a nevoilor din creier sunt conştiente, ci că *contează dacă o nevoie este simţită sau nu.* E posibil ca undeva, în fundal, proporţia apă versus sare din corp să fluctueze tot timpul, dar, când o simţi, vrei să bei. E posibil să fii în pericol în mod obiectiv fără să-ţi dai seama, dar, când simţi asta, cauţi modalităţi de scăpare.

Pentru lucruri diferite e nevoie de nume diferite, iar diferenţa dintre nevoile resimţite şi cele neresimţite face necesară introducerea unei distincţii terminologice. „Nevoile” sunt diferite de „afecte”. Nevoile corporale pot fi înregistrate şi reglate *autonom*, ca în exemplele controlului cardiovascular şi respirator, termoreglării şi metabolizării glucozei. Acestea se numesc funcţii „vegetative”, şi pe bună dreptate: nu au nimic conştient. De unde şi termenul „reflex” autonom. Conştiinţa intră în ecuaţie numai atunci când nevoile sunt simţite – momentul în care *îţi* solicită un travaliu. (Notă: pulsiunile evaluează solicitările de travaliu adresate *psihicului;* unele nevoi nu duc niciodată la acţiune voluntară şi, de aceea, nu devin niciodată conştiente. Tensiunea arterială este un exemplu notoriu din punct de vedere clinic; nu ştii nimic despre valorile ei prea mici sau prea mari până ce este deja prea târziu.)

Şi nevoile emoţionale pot fi gestionate automat, prin stereotipuri comportamentale precum „instinctele” (strategii de supravieţuire şi de reproducere înnăscute,

pe care Freud le-a plasat în centrul conceptului său de psihic inconştient). Dar nevoile emoţionale sunt de obicei mai dificil de satisfăcut decât cele corporale, din raţiuni pe care le voi explica în capitolul de faţă. Acesta este motivul pentru care senzaţiile pe care le evocă durează de regulă mai mult. O senzaţie dispare din conştiinţă atunci când nevoia pe care o anunţă a fost satisfăcută.

Al treilea lucru de remarcat în legătură cu senzaţiile este că nevoile resimţite sunt *prioritizate* în detrimentul celor neresimţite. Suntem permanent asaltaţi de numeroase nevoi. Funcţiile vegetative precum echilibrul energetic, controlul respirator, digestia, termoreglarea şi altele asemenea acţionează constant, la fel şi comportamentele stereotipe de tipuri diferite. Nu ai putea simţi toate aceste lucruri în acelaşi timp, în special pentru că poţi *face* câte un singur lucru (sau foarte puţine lucruri) pe rând. Trebuie făcută o selecţie. Ea are loc în funcţie de *context.* Priorităţile se stabilesc în funcţie de *r r r* intensitatea *relativă* a nevoilor (amploarea semnalelor de eroare) *în raport cu* oportunităţile existente în circumstanţele curente. Iată un exemplu simplu. În timp ce ţin o prelegere, vezica urinară mi se umple treptat, dar nu simt presiunea tot mai mare până ce îmi închei discursul, moment în care simt brusc nevoia să urinez. *Devin conştient* de semnalul de eroare datorită schimbării contextuale. În acel moment, simţind ocazia, nevoia mea corporală devine conştientă sub formă de senzaţie[6].

Prioritizarea nevoilor în acest fel are consecinţe importante. Cea mai importantă este că, *atunci când*

devii conştient de o nevoie, când o resimţi, ea îţi domină comportamentul voluntar.

Ce înseamnă „voluntar”? E opusul lui „automat”. Înseamnă supus unor *alegeri* ce ţin de aici-şi-acum. Alegerile pot fi făcute numai dacă se bazează pe un sistem de valori – ceea ce determină „calitatea de a fi bun” versus „calitatea de a fi rău”. Altfel, răspunsurile noastre la evenimente nefamiliare ar fi aleatorii. Iar aici închidem cercul, revenind la trăsătura fundamentală a afectului: valenţa. Tu decizi ce faci şi ce nu faci în baza *consecinţelor resimţite în urma acţiunilor tale.* Aceasta *ff r* este legea afectului. Astfel, comportamentul voluntar, ghidat de afect, are un enorm avantaj adaptativ faţă de comportamentul involuntar: ne eliberează de lanţurile automatismului şi ne permite să supravieţuim în situaţii neprevăzute.

Faptul că comportamentul voluntar trebuie să fie conştient relevă cea mai profundă funcţie biologică a senzaţiei: ne ghidează comportamentul în condiţii de *incertitudine.* Ne permite să determinăm în toiul acţiunii dacă un mod de a acţiona este mai bun sau mai prost decât altul. În exemplul foamei de aer, reglarea gazelor sangvine devine conştientă atunci când nu ai o soluţie de-a gata ca să te menţii în limitele tale viabile din punct de vedere fiziologic. În graba de a scăpa din încăperea plină cu dioxid de carbon, de pildă, cum ştii încotro s-o apuci? Nu ai mai fost niciodată în această situaţie (în nicio clădire în flăcări, ca să nu mai vorbim de exact această clădire), aşa că nu ai cum să anticipezi ce e de făcut. Acum trebuie să te hotărăşti dacă o iei pe aici sau pe acolo, în sus sau în jos etc. Iei asemenea

decizii *în funcţie de ceea ce simţi în timp ce cauţi o soluţie la problemă:* senzaţia de sufocare se intensifică sau se diminuează, în funcţie de direcţia în care te îndrepţi, cea greşită sau cea bună – adică în funcţie de creşterea sau de scăderea nivelului de oxigen din aer.

Senzaţia conştientă a alarmei de sufocare implică diverse structuri neuronale şi substanţe chimice dintre cele responsabile pentru controlul respirator inconştient, la fel cum senzaţia de foame selectează diferite sisteme cerebrale dintre cele responsabile cu reglarea autonomă a echilibrului energetic. Ştiinţa nu ar fi descoperit niciodată aceste lucruri dacă ar fi continuat să ignore senzaţiile.

Sunt mult mai multe de spus despre senzaţii, cum ar fi modul în care facilitează *învăţarea din experienţă* (prin legea afectului) şi modul în care sunt legate de gândire, dar ideile pe care tocmai le-am exprimat, pe scurt, explică *de ce* simţim. Am descris ce adaugă ceea ce simţim la repertoriul mecanismelor pe care noi, fiinţele vii, le folosim pentru a rămâne în viaţă şi a ne reproduce. Aceasta este contribuţia psihologiei la biologie. Selecţia naturală a influenţat aceste mecanisme de supravieţuire, dar, odată ce s-au dezvoltat senzaţiile afective – adică abilitatea noastră unică de organisme complexe de *a lua notă de propriile stări* –, în univers a apărut ceva cu totul nou: fiinţa subiectivă.

E greu de imaginat cum ai putea să ai o durere fizică fără să o simţi: să ai o durere pur şi simplu înseamnă să o simţi. Dar ce putem spune despre stările afective mai

subtile pe care le numim emoţii şi dispoziţii? Trebuie să simţi că eşti fericit ca să fii fericit? Oare nu e ceva obişnuit să realizezi prea târziu că eşti stăpânit de o dispoziţie proastă – să ajungi la concluzia că probabil eşti deprimat şi aşa mai departe?

Am început să explorăm afectul cu ajutorul formelor sale corporale, pentru că acestea sunt exemplele cele mai simple şi fără îndoială că au fost şi primele care au apărut în cursul evoluţiei. Consider că „zorii conştiinţei” nu au presupus ceva mai elaborat decât senzaţii somatice cu valenţe hedonice. Acum vreau să arăt că emoţiile umane sunt versiuni complexe ale unor asemenea lucruri. Şi ele sunt, în cele din urmă, semnale „de eroare” ce iau notă de devieri de la stările tale biologice preferate şi îţi spun dacă măsurile pe care le iei îţi îmbunătăţesc sau îţi înrăutăţesc situaţia.

Din păcate, nu există o clasificare a afectelor general acceptată în neuropsihologia curentă. Am făcut o distincţie între afectele corporale şi cele emoţionale, dar asemenea delimitări precise nu există în natură. Fac această delimitare pornind de la taxonomia lui Jaak Panksepp, care este larg – dar nu unanim – acceptată. El a împărţit şi mai departe afectele corporale, diversitatea lor fiind foarte mare, în subtipurile interoceptive („homeostatice”) şi exteroceptive („senzoriale”). Foamea şi setea, de exemplu, sunt afecte homeostatice, iar durerea, surpriza şi dezgustul sunt afecte senzoriale*[1]. Aşadar, pentru a fi clar, conform lui Panksepp există trei tipuri de afecte: homeostatice şi senzoriale (ambele tipuri fiind corporale) şi emoţionale

(care implică trupul, dar nu pot fi descrise drept „corporale” în niciun sens simplu). Gândeşte-te, de exemplu, că ţi-e dor de fratele tău, iar asta e o stare emoţională; nu este corporală în acelaşi fel în care sunt foamea şi durerea.

Panksepp şi-a fundamentat taxonomia pe studii asupra *stimulării cerebrale profunde* pe care el, studenţii lui şi mulţi alţii înaintea lor le-au efectuat pe mii de animale. I-am vizitat de multe ori laboratorul, care era o veritabilă menajerie, uneori plin de porumbei, alteori plin de pui de găină, câini, cobai, şobolani sau şoareci-de-câmp. (E posibil ca boala care l-a ucis în cele din urmă să fi fost cauzată de expunerea în exces la unele dintre aceste animale, mai ales la păsări.)

Nu încape îndoială că Jaak a simţit compasiune, chiar dragoste faţă de aceste animale. Dar trebuie să recunoaştem şi că au fost sacrificate pe altarul ştiinţei, în număr mare, fără ca ele să-şi aleagă soarta. E o ironie tristă că datorăm experimentelor lui Panksepp informaţia aproape sigură că animalele enumerate mai sus sunt creaturi capabile să simtă, cu emoţii intense, care, în esenţă, nu sunt diferite de cele resimţite de oameni. Ca urmare a acestor descoperiri şi a preocupărilor tot mai mari pentru problemele de etică ce decurg din ele, Panksepp şi-a petrecut ultimele decenii de viaţă studiind numai emoţiile pozitive.

În cele ce urmează, când trec de la observaţii despre animale la observaţii despre oameni şi invers, o fac în mod deliberat. Când colegii l-au acuzat de antropomorfism faţă de animale, Panksepp a spus că

mai degrabă s-ar recunoaşte vinovat de zoomorfism faţă de oameni. Obiectivul experimentelor sale a fost să identifice structurile şi circuitele cerebrale care activează în mod cert aceleaşi răspunsuri afective, nu numai la indivizi, ci şi la mai multe specii. Când e vorba de afecte emoţionale, s-a dovedit că şapte dintre ele pot fi reproduse nu numai la toate primatele, ci chiar *la toate mamiferele,* prin stimularea exact aceloraşi elemente chimice şi structuri cerebrale. (Multe dintre ele pot fi evocate şi la păsări, iar unele la toate vertebratele.) Mamiferele s-au separat de păsări acum aproximativ 200 de milioane de ani, iată cât de vechi sunt aceste emoţii. Totuşi, întrucât oamenii sunt mamifere, în continuare mă voi axa pe aceste şapte tipuri. Din câte ştim, ele sunt ingredientele *de bază* ale întregului repertoriu emoţional uman. Numeroasele noastre bucurii şi tristeţi par să fie outputurile acestor şapte sisteme, amestecându-se între ele şi cu procese cognitive superioare.

Probabil cea mai cunoscută taxonomie alternativă a „emoţiilor de bază” (aşa cum sunt numite) este cea elaborată de Paul Ekman[7]. Diferenţele apar în principal din faptul că Ekman a folosit o metodă diferită de cea a lui Panksepp, şi anume studiul expresiilor faciale şi al comportamentelor asociate. După cum a observat Charles Darwin cu mult timp în urmă în *Expresia emoţiilor la om şi animale* (1872), acestea au elemente comune remarcabile la diverse specii de mamifere. Totuşi, deşi Panksepp şi Ekman clasifică afectele în mod diferit (de exemplu, Ekman consideră că dezgustul este un afect „emoţional”, iar Panksepp îl consideră

„senzorial”), există un larg acord în privinţa afectelor. Nimeni nu se îndoieşte că dezgustul există, aşa că, într-un anumit sens, nu prea contează cum îl clasificăm.

Principala opinie diferită este cea a Lisei Feldman Barrett. Din nou, dezacordul poate fi pus în primul rând pe seama unor diferenţe metodologice[8]. Ea se axează pe emoţiile autoraportate la oameni şi, deloc surprinzător, descoperă că există variaţii considerabile în ce priveşte modul în care diverşi oameni (şi diverse culturi) caracterizează şi analizează sentimentele. Asta nu contrazice faptul că tipurile naturale de bază se află sub suprafaţa construită social. Voi ilustra imediat mecanismele prin care apare această variabilitate, dar explicaţia pe scurt este: reflexele şi instinctele sunt instrumente nerafinate, dar funcţionale pentru supravieţuire şi succes reproductiv, dar ele nu ne pot echipa adecvat pentru multitudinea de situaţii imprevizibile şi medii în care ajungem. Prin urmare, trebuie să completăm în mod adaptativ răspunsurile înnăscute cu învăţarea din experienţă. Faptul că oamenii fac acest lucru cu atâta uşurinţă e principalul motiv pentru care, până la urmă, am ajuns să dominăm lumea într-o asemenea măsură.

Programele instinctive aflate la baza acţiunilor oamenilor sunt de regulă atât de condiţionate prin învăţare, încât nu mai pot fi recunoscute drept „instinctive”. Totuşi, instinctele şi reflexele sunt întotdeauna în fundal. Întreaga teorie psihanalitică se întemeiază pe ideea că, dacă îţi baţi capul să le găseşti, tendinţele instinctive implicite pot fi identificate întotdeauna în spatele intenţiilor explicite.

Iar acum, o prezentare rapidă a emoţiilor instinctive, conform taxonomiei lui Panksepp[9]. Diverşi oameni de ştiinţă clasifică afectele în mod diferit, aşa că folosesc şi cuvinte diferite. Panksepp a scris cu majuscule denumirile emoţiilor de bază pentru a le deosebi de utilizarea colocvială – adică pentru a arăta că vorbeşte despre funcţiile cerebrale sub toate aspectele lor, nu numai despre sentimente.

(1) DORINŢA SEXUALĂ. Nu suntem permanent excitaţi sexual. Sentimentul erotic intră în conştiinţă numai atunci când sexul este prioritizat în faţa altor motivaţii, iar asta se întâmplă în contextul unor nevoi şi oportunităţi fluctuante. Când dorinţa sexuală este activată, o simţi, apoi sentimentele erotice îţi dirijează acţiunile. Când eşti excitat sexual, acorzi atenţie altor detalii decât atunci când îţi este frică, de exemplu, şi te şi comporţi diferit. Astfel, conştiinţa exteroceptivă şi comportamentul tău voluntar sunt determinate de starea ta interioară; resimţi lumea în mod diferit – literalmente aducându-ţi pe cap experienţe diferite –, în funcţie de ceea ce simţi. Acesta este şi motivul pentru care nu poţi simţi în acelaşi timp excitare sexuală şi repulsie motivată de frică, nu le poţi prioritiza pe amândouă. Când nevoia de siguranţă este prioritizată, mintea îţi anulează motivaţiile sexuale.

Nu e clar dacă DORINŢA SEXUALĂ ar trebui clasificată ca afect „corporal” sau „emoţional”. Unii se îndoiesc chiar că sexualitatea este o nevoie. Acesta este un exemplu excelent al diferenţei dintre nevoi (inconştiente) şi afectele (conştiente) pe care le

generează. Când ne implicăm în acte sexuale, de obicei nu încercăm să ne îndeplinim datoria biologică. De fapt, foarte frecvent, sperăm să *nu* ne reproducem. Ca în cazul gustului dulce versus rezervele de energie, ceea ce ne motivează ca fiinţe subiective este căutarea plăcerii erotice, şi nu a succesului reproductiv. Adică suntem conduşi de *senzaţii.* Dar organismele vii *au nevoie* să se reproducă, cel puţin în general. Iată principalul motiv pentru care sexul a devenit plăcut din punct de vedere subiectiv, prin selecţie naturală.

Am spus „în general” deoarece nu întreaga activitate sexuală duce la reproducere, ci numai suficient din ea cât să existe specia. Acest lucru ilustrează alt principiu central: utilitatea limitată a comportamentelor înnăscute pentru a ne satisface nevoile emoţionale. În actul sexual, aspectele înnăscute se reduc doar la congestie şi lubrifiere genitală, lordoză (arcuirea spatelui, care face vaginul disponibil pentru penetrare), încălecare, intromisiune, împingere şi ejaculare. Împreună cu aceste reflexe, atingerea clitorisului sau a penisului (care sunt organe echivalente) într-un anumit ritm produce senzaţii plăcute care anticipează eliberarea tensiunii sexuale prin orgasm, în cele din urmă, până la satisfacere. Aceste mecanisme involuntare nu ne echipează pentru misiunea dificilă de a-i convinge pe alţi oameni – mai ales pe cei de care ne simţim atraşi – să se supună dorinţei noastre de a face sex cu ei. Motivul principal pentru care nevoile „emoţionale” sunt mai dificil de satisfăcut decât cele corporale este că, de regulă, ele implică alţi agenţi capabili de a simţi, care au propriile nevoi; ele nu sunt

simple substanţe precum mâncarea şi apa. Prin urmare, pentru a ne satisface nevoile sexuale, trebuie să ne suplimentăm cunoaşterea înnăscută cu alte abilităţi, dobândite prin *învăţare.* Numai astfel se explică marea diversitate a activităţilor noastre sexuale, alături de forma „medie” moştenită prin selecţie naturală.

Să observăm că învăţarea nu *şterge* reflexele şi instinctele; le elaborează, le completează şi le domină, dar ele există în continuare. Felinarele stradale ilumineză drumurile noaptea, dar nu pot să elimine complet întunericul. Mecanismul obişnuit de actualizare a memoriei pe termen lung, „reconsolidarea” (pe care îl descriu în capitolul 10), nu se aplică în cazul reflexelor şi instinctelor. Iar asta pentru că reflexele şi instinctele nu sunt *amintiri,* ci înclinaţii fundamentale „determinate genetic” în fiecare specie prin selecţie naturală.

Varietatea comportamentelor noastre sexuale se lărgeşte şi mai mult datorită faptului că circuitul cerebral al DORINŢEI SEXUALE tipice indivizilor de sex feminin şi al celei tipice indivizilor de sex masculin există la fiecare mamifer. Tendinţa care ajunge să domine este determinată de factori diverşi, inclusiv de evenimente genetice şi intrauterine[10]. Nu voi intra aici în detalii de anatomie şi de chimie, subliniez doar că ambele circuite apar în hipotalamus şi se termină în substanţa cenuşie periapeductală, abreviată PAG. (Am arătat localizarea PAG în figura 6. Aflăm imediat de ce este atât de importantă.) Cu alte cuvinte, ele sunt în întregime subcorticale—.

(2) CĂUTAREA. Toate nevoile corporale (şi cele sexuale) – care sunt înregistrate de „detectoarele de nevoi” localizate în principal în hipotalamusul medial – activează această a doua pulsiune emoţională, pe care am prezentat-o în capitolul 1. Este aproape sinonimă cu conceptul de „libido” al lui Freud, dar Freud nu ştia că DORINŢA SEXUALĂ doar *activează* acest sistem; DORINŢA SEXUALĂ şi CĂUTAREA nu sunt acelaşi lucru. CĂUTAREA generează un comportament explorator de „căutare a hranei”, însoţit de o stare emoţională conştientă ce poate fi caracterizată drept expectaţie, interes, curiozitate, entuziasm sau optimism. Să ne gândim la un câine în aer liber: indiferent care îi sunt nevoile corporale curente, căutarea hranei îl impulsionează să interacţioneze pozitiv cu mediul, astfel încât să şi le poată satisface. Aproape tot ce avem nevoie ca fiinţe vii se găseşte „undeva afară”. Prin căutarea hranei învăţăm, aproape accidental, ce lucruri din lume ne satisfac fiecare dintre dorinţe. Astfel, codificăm relaţiile lor de tip cauză-efect. Acest lucru ilustrează din nou modul în care instinctele stereotipe duc la învăţare individualizată.

CĂUTAREA este neobişnuită printre emoţiile de bază prin faptul că acceptă incertitudinea *în mod proactiv.* Aceasta este originea comportamentelor de căutare a noutăţii şi chiar de asumare a riscului. Căutarea hranei ne face să explorăm lucruri interesante, astfel încât să ştim la ce să ne aşteptăm când le vom mai întâlni. De exemplu, odată ce un câine a explorat un gard viu şi s-a familiarizat cu conţinutul lui, va fi mai puţin interesat de el data viitoare. În consecinţă, CĂUTAREA este emoţia

noastră „implicită”. Atunci când nu suntem dominaţi de vreunul dintre celelalte afecte („legate de sarcini”), conştiinţa noastră tinde spre acest sens generalizat al interesului faţă de lume.

Din punct de vedere anatomic, neuronii circuitului CĂUTĂRII apar din aria tegmentală ventrală a trunchiului cerebral, de unde urmează un curs ascendent, prin hipotalamusul lateral spre *nucleus accumbens,* amigdală şi cortexul frontal (vezi figura 2). Din punct de vedere chimic, neuromodulatorul care îl comandă este dopamina, „materia viselor” (vezi capitolul 1)[12]. Asta ne spune ceva interesant despre CĂUTARE, şi anume că poată fi activată chiar şi în timpul somnului, prin solicitări de travaliu adresate psihicului, ducând la activităţi de rezolvare a problemelor ce trebuie ghidate de sentimente conştiente. De aici, visele.

(3) FURIA. Deşi abordăm mediul într-un mod pozitiv prin CĂUTARE, cu credinţa optimistă că ne va satisface nevoile, lucrurile nu merg întotdeauna bine pentru noi. Aşa cum preistoria evoluţiei ne-a dotat cu reflexe şi instincte ce anticipează destul de bine căile pe care ne putem satisface nevoile corporale, tot astfel ne naştem cu tendinţe emoţionale ce anticipează căi pe care putem ieşi din încurcătură. În situaţii complicate cu semnificaţie universală, sentimentele adecvate sunt prioritizate pentru a domina comportamentul. Astfel suntem scutiţi de costurile biologice ale necesităţii de a reinventa roţile care le-au permis strămoşilor noştri să supravieţuiască şi să se reproducă. Emoţiile sunt o

moştenire preţioasă. Ele transmit abilităţi de supravieţuire înnăscute – cunoaştere inconştientă, implicită – sub forma conştientă a senzaţiilor care ne pot ghida acţiunile în mod explicit.

Când se declanşează sistemul FURIEI – aşa cum se întâmplă când intervine ceva între noi şi orice ne-ar putea satisface nevoile curente –, conştiinţa este definită de sentimente ce variază de la frustrare iritată la furie oarbă. Reflexele şi instinctele ce se manifestă în asemenea situaţii includ piloerecţia (erecţia părului), protruzia ghearelor, sâsâiturile, mârâielile şi arătarea colţilor, urmate de „atacul afectiv": te arunci asupra ţintei mâniei tale şi o muşti, o împingi sau o loveşti până când se potoleşte.

De ce *simţi* afectele care însoţesc asemenea comportamente? Răspunsul este ca şi mai sus: sentimentele îţi spun cum te descurci, dacă lucrurile merg bine sau prost, când încerci să scapi de un obstacol – care, adesea, încearcă să scape de tine în acelaşi timp. Simţi gustul dulce al victoriei sau amarul înfrângerii. Acest lucru ghidează ce faci în continuare, inclusiv posibilitatea ca durerea (un afect senzorial suprimat în timpul atacului afectiv) să fie prioritizată, înlocuind astfel FURIA şi punând capăt luptei – şi poate ducând la fugă.

Cum ar putea toate acestea să se desfăşoare automat, fără o evaluare conştientă continuă? Întrebarea se aplică cu atât mai mult rolului afectului în *gândire* – subiect pe care l-am putea introduce cu folos acum. Gândirea înseamnă acţiune „virtuală", capacitatea de a testa lucruri în imaginaţie, capacitate

care, din raţiuni biologice evidente, salvează vieţi. Această capacitate nu este tipică oamenilor, dar în cazul nostru a ajuns la un nivel superior de dezvoltare. Să apelăm la un caz concret. Imaginează-ţi situaţia următoare, derivată din experienţa mea. Directorul mă ceartă în biroul lui. Mă simt tot mai furios. Răspunsul instinctiv este atacul afectiv. Acum *mă gândesc* la consecinţele posibile. Aşa că, în loc să mă arunc asupra lui, îmi reprim tendinţa instinctivă spre acţiune şi îmi *imaginez* seria de alternative; îmi caut calea de ieşire în funcţie de ceea ce simt. În cele din urmă, mă opresc la o soluţie satisfăcătoare: după ce ies din birou, când nu mă vede nimeni, îi sparg cauciucurile de la maşină. În felul acesta, îmi reduc FURIA fără să sufăr consecinţe cumplite. Vedem astfel, încă o dată, de ce stereotipurile comportamentale înnăscute trebuie să fie completate de învăţarea din experienţă, inclusiv de forma imaginară a experienţei numită gândire. Când ne confruntăm cu frustrări reale, ce includ frecvent nevoi *conflictuale* (în acest caz, FURIE versus FRICĂ), soluţiile instinctive nu sunt suficiente. Dar de notat, din nou: dacă răspunsurile instinctive sunt completate de învăţare nu înseamnă că ele sunt eliminate. Eu m-am hotărât să nu-l atac pe directorul meu, dar înclinaţia să o fac mi-a rămas şi va apărea din nou în situaţii viitoare similare. (Acest exemplu nu e valabil numai în cazul oamenilor. Un câine poate că nu ar fi ajuns la o asemenea soluţie, dar primatele fac tot felul de năzbâtii[13].)

Emoţiile precum FURIA nu sunt „simple” sentimente. Emoţiile joacă un rol fundamental în supravieţuire. Să ne imaginăm consecinţele dacă nu am avea pretenţii la

resursele disponibile şi nu i-am împiedica pe alţii să-şi însuşească partea noastră. Dacă nu am putea deveni frustraţi, iritaţi sau furioşi, nu am fi înclinaţi să luptăm pentru ce ne trebuie, iar în acest caz, mai devreme sau mai târziu, am muri. E uşor să trecem cu vederea funcţia biologică a emoţiei în condiţiile civilizate în care trăim astăzi. Dar trăim astfel (în aşezări permanente cu legi artificiale ce reglementează comportamentul social) de doar aproximativ 12.000 de ani. Civilizaţia este o caracteristică recentă a existenţei mamiferelor, nu a jucat niciun rol în modelarea creierilor noştri.

Gândirea conştientă necesită cortex, dar nu şi sentimentele care o ghidează. Circuitul care mediază FURIA este aproape în întregime subcortical şi, ca orice alte circuite afective, are destinaţia finală în PAG din trunchiul cerebral[14].

(4) FRICA. Dicotomia luptă/fugă arată că atacul afectiv nu este întotdeauna cea mai bună metodă de a-i face faţă unui adversar. Factorii contextuali ce separă lupta de fugă sunt înscrişi în amigdală, care mediază şi FURIA, şi FRICA[15].

Majoritatea mamiferelor „ştiu” din prima zi că unele lucruri sunt inerent înfricoşătoare. Puii nou-născuţi ai rozătoarelor, de exemplu, încremenesc când sunt expuşi chiar la un singur fir de păr de pisică – deşi nu au avut niciodată de-a face cu pisicile şi nu ştiu nimic despre atitudinea lor faţă de şoareci. E uşor să înţelegem de ce stau lucrurile aşa: dacă fiecare şoarece ar trebui să înveţe din experienţă cum să reacţioneze la pisici, atunci şoarecii ar dispărea ca specie. Din nou, vedem enorma valoare biologică a emoţiilor.

Noi, oamenii, ne temem de pericole precum înălţimile, locurile întunecate şi creaturile care se târăsc spre noi şi le evităm cu ajutorul aceloraşi instincte şi reflexe ca ale altor mamifere: prin comportamente de încremenire şi de fugă. Spre deosebire de reflexul vasovagal de leşin discutat mai sus, aceste comportamente de „evadare” sunt facilitate de respiraţia rapidă, de ritmul cardiac intensificat şi de redirecţionarea sângelui din intestine spre musculatura scheletului. (De unde şi pierderea controlului sfincterelor asociată cu frica extremă.) Ca în cazul altor emoţii, sentimentul conştient de frică îţi spune dacă te îndrepţi spre siguranţă sau dacă te îndepărtezi de ea.

Un exemplu interesant este oferit de pacienta S.M., care avea aproape 30 de ani când a fost publicat cazul ei pentru prima dată. Suferea de sindromul Urbach-Wiethe, o afecţiune genetică rară ce duce la calcificarea bilaterală a amigdalei. Nu simţea teama. În consecinţă:

> [Ea] a fost victima a numeroase infracţiuni şi incidente traumatice care i-au pus viaţa în pericol. A fost ameninţată cu cuţitul şi cu arma, a fost aproape omorâtă într-un incident de violenţă domestică şi a primit ameninţări explicite cu moartea în numeroase ocazii. În pofida faptului că multe dintre aceste situaţii îi puneau viaţa în pericol, S.M. nu manifesta niciun semn de disperare, urgenţă sau alte răspunsuri comportamentale ce ar fi asociate în mod normal cu asemenea incidente. Numărul disproporţionat de mare de evenimente traumatice din viaţa lui S.M. este pus pe seama [...] incapacităţii sale severe de a detecta ameninţările iminente din

mediu şi de a învăţa să se ţină departe de situaţiile cu potenţial de periculozitate[16].

Eu am examinat un mare număr de asemenea pacienţi, întrucât sindromul Urbach-Wiethe are o incidenţă neobişnuit de mare într-un colţ izolat al Africii de Sud numit Namaqualand, care se întâmplă să fie în apropiere de locul unde m-am născut. (Gena defectă a fost adusă acolo de un colonist german, apoi s-a concentrat într-o comunitate izolată.) Mai ales visele lor mi s-au părut interesante – sunt scurte, simple şi pline de aspiraţii. Una dintre pacientele pe care le-am studiat, al cărei soţ era şomer, a visat: „Soţul meu şi-a găsit un loc de muncă. Eram foarte fericită”. Alta, mamă a unui copil cu dizabilităţi, a visat: „Fata mea putea să meargă. Eram foarte fericită”. Un pacient al cărui tată murise a visat: „L-am văzut pe tata din nou. Eram aşa de fericit!”. Aceste vise sunt tipice pentru cei care suferă de sindromul Urbach-Wiethe, a căror imaginaţie neînfricată nu se aşteaptă la niciun pericol în procesul satisfacerii dorinţelor lor[17].

Majoritatea oamenilor par că se nasc cu anumiţi declanşatori specifici ai FRICII. Oare cum ar fi dacă fiecare dintre noi ar trebui să înveţe din experienţă ce se întâmplă când sari de pe o stâncă sau când pui mâna pe o viperă? Acesta este motivul pentru care ne tragem din strămoşi care nu se simţeau înclinaţi să încerce. Cei care au încercat nu sunt strămoşii noştri pentru că ei nu au avut urmaşi. Avem toate motivele să fim recunoscători pentru această moştenire.

Dar trebuie să învăţăm de ce *altceva* să ne temem.

Învăţăm din experienţă – inclusiv cu ajutorul gândirii – că un rău asemănător ne poate fi cauzat şi de alte lucruri în afară de căderea de la o înălţime ameţitoare şi de muşcăturile de şarpe. Evoluţia nu ar fi avut cum să anticipeze prizele electrice şi şocul pe care îl produc, de exemplu, dar ele sunt la fel de periculoase ca viperele. De asemenea, trebuie să învăţăm ce altceva *să facem* când ne este frică, pentru a ne completa răspunsurile instinctive. Nu este adaptare să încremeneşti în faţa a ceva care te sperie sau să fugi de acel ceva, tot aşa cum nu este adaptare să ataci pe oricine te frustrează. Ar trebui să fie deja clar ce rol joacă sentimentele conştiente în procesul învăţării – ne spun ce funcţionează şi ce nu înainte să fie prea târziu şi, astfel, ne ajută să rămânem în viaţă.

Condiţionarea prin frică relevă alte lucruri importante despre ce este conştient şi ce nu. Una dintre trăsăturile ei speciale este „învăţarea printr-o singură expunere” – nu trebuie decât să-ţi bagi degetul într-o priză *o singură dată* ca să nu mai repeţi acţiunea. E uşor să înţelegem de ce: ai avut noroc că ai supravieţuit prima dată, aşa că de ce să repeţi experienţa? Totuşi, la fel ca în cazul tuturor celorlalte mecanisme biologice din spatele emoţiilor, nu e nevoie să cunoşti acest lucru într-un sens „declarativ” – condiţionarea se întâmplă pur şi simplu automat, iar asta deoarece condiţionarea prin FRICĂ nu necesită implicarea cortexului. Ea poate să aibă loc în prima copilărie, chiar înainte de maturizarea hipocampului (structura corticală responsabilă cu fixarea amintirilor declarative pe termen lung). Din acest motiv, ca în cazul lui Claparède,

multor oameni sănătoşi din punct de vedere neurologic le este frică de anumite lucruri fără să ştie de ce.

Specialiştii în cogniţie atribuie acest lucru învăţării „inconştiente”, dar, după cum am spus deja, o fac doar pentru că neglijează afectul. E adevărat că mulţi oameni nu conştientizează motivele pentru care se tem de diverse lucruri, dar sunt foarte conştienţi de sentimentele asociate. Trebuie doar să simţi pentru a ghida comportamentul voluntar. Să ne amintim de experimentul cu succesiunea rapidă a unor cuvinte, descris mai sus: dacă termeni precum „criminal” şi „violator” sunt asociaţi subliminal cu chipul A, iar „grijuliu” şi „generos” cu chipul B, subiecţii cercetării vor simţi o preferinţă când li se va cere, mai târziu, să facă o alegere, deşi nu pot spune de ce. Această alegere e ghidată de „instinct”, dar sentimentele rămân cu uşurinţă nerecunoscute, aşa că sunt descrise din punct de vedere cognitiv prin cuvinte precum „a ghici”[18].

Astfel se explică o mare parte dintre nelămuririle din jurul „emoţiilor inconştiente” din ştiinţa cognitivă. Nu atât *emoţiile* sunt inconştiente, cât lucrurile *cognitive* la care se referă. După cum am văzut mai sus în legătură cu gândirea, cu siguranţă ajută să ştii la ce se referă sentimentele tale, dar această cunoaştere nu este esenţială. De fapt, uneori este mai bine să *nu* gândeşti înainte să acţionezi, nu în ultimul rând pentru că gândirea presupune timp.

Aceleaşi afirmaţii sunt valabile şi în cazul condiţionării prin frică. Odată ce ai învăţat să te temi de ceva – mai ales dacă nu ştii în mod conştient motivul –, asocierea este aproape ireversibilă. În memorabila

formulare a lui Joseph LeDoux, amintirile legate de frică sunt „de neuitat”[19]. Asta ne spune lucruri importante despre memoria inconştientă în general, pe care le voi rezuma mai jos. Deocamdată, menţionez doar că amintirile „nondeclarative” (precum cele emoţionale şi procedurale) sunt greu de uitat, din acelaşi motiv pentru care sunt inconştiente: conţin mai puţină incertitudine (adică sunt mai generalizabile) şi, prin urmare, mai puţin susceptibile de revizuire contextuală. Acesta este modul în care comportamentele dobândite devin stereotipe şi automate. În măsura în care obiectivul cunoaşterii este să înveţi cum să-ţi satisfaci nevoile în lume, automatizarea este idealul învăţării.

(5)PANICA/MÂHNIREA. Anxietatea de separare este diferită de FRICĂ. Apare iniţial în fragedă copilărie, când devii ataşat instinctiv de mama ta (sau de principalul tău îngrijitor). Spre deosebire de condiţionarea prin frică, dar din motive de natură biologică la fel de întemeiate, acest proces durează aproximativ şase luni: o singură situaţie de îngrijire afectuoasă nu este suficientă pentru a arăta că te poţi baza pe o anumită persoană pentru totdeauna.

Când mamiferele sunt separate de fiinţele de care s-au ataşat, se desfăşoară o secvenţă stereotipă, începând cu un comportament de „protest”, urmat de „disperare”. Etapa protestului este caracterizată de senzaţii de panică, însoţită de vocalizări ale detresei şi comportamente de căutare. Panica se combină frecvent cu furia, care evocă alt conflict, de data asta între

PANICĂ/MÂHNIRE şi FURIE. Prima emoţie te face să vrei să-ţi ţii îngrijitorul aproape pentru totdeauna, pe când cealaltă te face să vrei să-l distrugi. Vinovăţia, o emoţie *secundară* care inhibă FURIA, este de obicei rezultatul învăţat. Acesta este un bun exemplu al modului în care emoţiile secundare (precum vinovăţia, ruşinea, invidia şi gelozia) apar din situaţii conflictuale. Spre deosebire de emoţiile de bază, acestea sunt constructe învăţate – hibrizi între emoţie şi cogniţie (după cum arată cercetarea lui Barrett).

Etapa disperării este caracterizată de sentimente de deznădejde – literalmente de „renunţare”. Explicaţia standard este: dacă plânsul şi căutările puiului separat nu duc repede la reuniune, atunci costurile potenţiale ale alertării prădătorilor cu privire la starea lui vulnerabilă încep să depăşească beneficiile. De asemenea, dacă puiul se îndepărtează prea mult de adăpost, scad şansele să-l găsească mama când se întoarce. Astfel, din punct de vedere statistic, cum s-a întâmplat cu sincopa vasovagală, renunţarea (deşi poate fi foarte dureroasă) devine strategia de supravieţuire moştenită.

Iată o descriere clasică a cascadei de separare la copii, făcută de psihanalistul John Bowlby:

> [Protestul…] poate începe imediat sau poate fi amânat; durează de la câteva ore la o săptămână sau mai mult. În acest timp, copilul mic pare grav tulburat de faptul că şi-a pierdut mama şi încearcă să o regăsească prin utilizarea deplină a resurselor sale limitate. Adesea plânge tare, îşi zgâlţâie pătuţul, se

aruncă în toate părţile şi se uită cu nerăbdare spre orice semn sau sunet care s-ar putea dovedi că este mama lui absentă. Întregul lui comportament sugerează o expectanţă puternică că ea se va întoarce. Între timp, tinde să respingă orice alte persoane care se oferă să facă ceva pentru el, deşi unii copii se vor agăţa cu disperare de o asistentă medicală.

[Disperarea...] urmează protestului, preocuparea copilului faţă de mama absentă este încă evidentă, deşi comportamentul lui sugerează tot mai multă deznădejde. Mişcările fizice active se diminuează sau încetează şi e posibil să plângă monoton sau intermitent. Copilul e retras şi inactiv, nu le cere nimic oamenilor din mediul său şi pare să fie într-o stare de doliu profund[20].

Bineînţeles, cea din urmă stare este similară depresiei, care este însoţită frecvent de vinovăţie. În consecinţă, Panksepp şi alţii (inclusiv eu) au aplicat această explicaţie a mecanismelor cerebrale ale PANICII/MÂHNIRII pentru a dezvolta tratamente noi ale tulburărilor de dispoziţie[21]. Din punct de vedere chimic, tranziţia de la „protest” la „disperare” este mediată de nişte peptide numite opioide, care întrerup dopamina (vezi efectele lor în cazul descris mai jos, p. 132). Din acest motiv, depresia este caracterizată de contrariile în oglindă ale senzaţiilor ce caracterizează CĂUTAREA[22]. Traiectoria anatomică a componentei PANICĂ a acestui sistem coboară din girusul cingular

anterior în PAG, locul unde se termină toate circuitele emoţiilor[23]. (Voi explica mai jos cum de toate ciclurile afective se termină *şi* încep în PAG din trunchiul cerebral.)

E interesant că acest circuit mediat de opioide a evoluat din sistemul analgezic mai vechi al creierului – suferinţa mintală a pierderii este o elaborare a mecanismelor corporale pentru durerea senzorială[24]. Acesta este un bun exemplu al tranziţiei unitare existente în natură între afectele senzoriale ce salvează viaţa şi trăirile emoţionale. Emoţiile nu au nimic „fictiv”. Sentimentele dureroase asociate cu separarea şi pierderea – cuplate cu învăţarea din experienţă – joacă un rol cauzal în asigurarea supravieţuirii mamiferelor şi păsărilor, care *au nevoie* de îngrijitori. Acest fapt e valabil şi după copilărie: circuitele cerebrale descrise mai sus mediază legăturile de ataşament de-a lungul vieţii, după cum mediază, din păcate, şi multe alte forme de dependenţă.

(6)GRIJA este cealaltă latură a ataşamentului – nu numai că avem nevoie noi înşine de grijă afectuoasă, ci trebuie să avem grijă şi de cei mici, mai ales de progeniturile noastre. Aşa- numitul instinct matern există în noi toţi, dar nu în acelaşi grad, deoarece este mediat de substanţe întâlnite la niveluri superioare (în general) la femei: estrogen, prolactină, progesteron şi oxitocină – toate crescând semnificativ în timpul sarcinii şi al naşterii[25]. Remarcabilă este şi suprapunerea dintre chimia cerebrală şi circuitele pentru GRIJĂ,

PANICĂ/MÂHNIRE şi DORINŢA SEXUALĂ tipică femeilor[26]. Chiar şi numai aceste fapte pot să explice de ce depresia este mult mai întâlnită (de aproape trei ori mai frecvent) la femei decât la bărbaţi. Aproximativ 80% dintre persoanele de sex feminin ştiu cumva din copilărie că este „bine” să-i legene pe copii înspre partea stângă a liniei mediane a corpului, iar bărbaţii tind să descopere (instinctiv) acest lucru după ce devin taţi[27]. Totodată, chiar şi băieţi complet neexperimentaţi ştiu ce să facă de obicei când plânge un bebeluş. Nu îl împung cu degetele şi nici nu îl ridică de un picior ca să vadă dacă ajută la ceva, ci ştiu (prezic în mod instinctiv) că este „bine” să-l ţină aproape şi să-l legene în timp ce scot sunete liniştitoare.

Şi totuşi, după cum învaţă fiecare părinte, aceste lucruri nu sunt suficiente. Ca să creşti cu succes un copil până devine adult, e nevoie de mult mai mult decât instinct. Aşadar, ca şi în cazul celorlalte emoţii, trebuie să învăţăm din experienţă ce este de făcut în numeroasele situaţii neprevăzute care apar. Ca şi în cazul celorlalte emoţii, deciziile noastre în acest sens sunt ghidate de sentimente (de grijă şi preocupare) – care ne spun dacă lucrurile merg bine sau rău. Un alt motiv pentru care pulsiunea de a îngriji nu e suficientă este că nu simţim *numai* dragoste faţă de copiii noştri, după cum va confirma orice părinte. Conflictele ce rezultă de aici trebuie rezolvate prin procese cognitiv-emoţionale hibride.

Învăţarea modului de a împăca flexibil diversele nevoi emoţionale stă la baza maturităţii şi sănătăţii mintale. De

exemplu, să ne gândim la relaţiile de cuplu de durată, care necesită integrarea judicioasă a DORINŢEI SEXUALE cu ataşamentul de tip PANICĂ/MÂHNIRE specific copiilor (să ne gândim la complexul virgină/prostituată, ce apare din incapacitatea de a împăca trăirile sexuale cu cele de afecţiune). Şi legăturile afective sunt dificil de împăcat cu sistemul dinamic al CĂUTĂRII (să ne gândim la fiorul pe care ni-l dă noutatea), nu doar frustrările inevitabile ce cauzează FURIA (de unde şi ubicuitatea conflictelor domestice), ceea ce intră în conflict cu preocupările GRIJII afectuoase ş.a.m.d. Menţinerea relaţiilor pe termen lung este doar un exemplu al numeroaselor provocări cu care se confruntă inima omului. Ca să gestionăm problemele vieţii, folosim emoţiile pe post de busolă. Ceea ce simţim ghidează orice învăţare din experienţă, în formele diferite pe care le-am prezentat. Dar biologia ne oferă în ajutor încă o pulsiune:

(7) JOACA. *Avem nevoie* să ne jucăm. Joaca este mediul prin care se revendică şi se apără teritorii, se formează ierarhii sociale şi se stabilesc şi se menţin graniţele în interiorul şi în afara grupului.

Oamenii sunt adesea surprinşi să afle că e o pulsiune biologică, dar toţi puii de mamifere se implică în joacă dinamică şi destul de dură. Dacă nu îşi iau porţia de joacă într-o zi, vor încerca să şi-o recupereze a doua zi. Ştim ce înseamnă încăierarea în joacă, deşi forma pe care o ia variază puţin de la o specie de mamifere la alta. O sesiune de joacă începe cu un anumit gest sau o anumită postură în semn de „invitaţie”, apoi, dacă este

acceptată, începe joaca. Animalul sau copilul îl fugăreşte pe celălalt plin de exuberanţă, apoi se opresc ca să se ia la trântă sau să se gâdile, fiecare stând deasupra, pe rând – toate acestea însoţite de hohote de râs sau de vocalizarea echivalentă, în funcţie de specie (chiar şi şobolanii „râd”)[28]. Apoi sar din nou în picioare, fugărindu-se în direcţia opusă. Trăirea emoţională asociată este la fel de universală: se numeşte distracţie.

Copiii iubesc joaca. Totuşi, empiric, majoritatea episoadelor de joacă se încheie cu lacrimi. Acest lucru oferă un indiciu important pentru a înţelege despre ce e vorba de fapt, dintr-o perspectivă biologică: e vorba de *găsirea limitelor* a ceea ce este tolerabil şi permisibil din punct de vedere social. Când jocul nu mai este distractiv pentru unul dintre participanţi, adesea din cauză că hotărăşte că nu eşti „corect”, acesta nu mai vrea să se joace. Şi-a atins limita. Marcarea unor asemenea limite este esenţială pentru formarea şi menţinerea unor grupuri sociale stabile. Iar supravieţuirea unui grup este importantă pentru supravieţuirea fiecărui membru al grupului la speciile sociale cum este a noastră.

În acest sens, un criteriu important este *dominarea.* În orice situaţie de joacă, unul dintre participanţi îşi asumă rolul de conducere, iar celălalt se supune. Aşa e distractiv pentru amândoi, atât timp cât cel dominant nu insistă să fie la comandă *tot* timpul. Raportul acceptabil al schimbării rolurilor pare să fie în jur de 60:40. Conform „regulii 60:40” a *reciprocităţii,* jucătorul supus continuă să joace atât timp cât i se oferă suficiente oportunităţi să preia conducerea.

Acest lucru relevă o a doua funcţie a JOCULUI: stabilirea unor ierarhii sociale. În consecinţă, încăierarea în joacă lasă locul (odată cu dezvoltarea) unor jocuri mai organizate şi deschis competitive. Desigur, joaca nu se limitează la simularea unor încăierări. Noi, oamenii, ne implicăm în diverse jocuri de rol, în care participanţii încearcă diferite roluri sociale (de exemplu, de-a mama şi copilul, profesorul şi elevul, doctorul şi pacientul, hoţul şi vardistul, cowboyul şi indianul[29], regele şi tâlharul – de observat omniprezentele ierarhii de status şi de putere). Nu ştim ce se petrece în imaginaţia altor mamifere când se joacă, dar putem presupune cu siguranţă că şi ele „încearcă" diferite roluri sociale, învăţând astfel ce este acceptabil şi ce este inacceptabil.

Acest lucru sugerează o a treia funcţie biologică pentru JOACĂ. Îţi cere (şi te condiţionează) să ţii cont de sentimentele altora. Dacă nu ţii cont, ceilalţi vor refuza să se joace cu tine, iar tu vei fi lipsit de plăcerea enormă pe care ţi-o aduce joaca. Deşi cel care face bullying acaparează toate jucăriile, el va fi privat de toată distracţia. Se pare că acesta e motivul pentru care JOACA a evoluat (şi pentru care este însoţită de atâta plăcere): promovează formaţiuni sociale viabile. Pe scurt, este un vehicul important pentru dezvoltarea *empatiei*[30].

Episoadele de joacă încetează abrupt atunci când îşi pierd calitatea de „ca şi cum". Dacă îţi încui surioara undeva şi arunci cheia, atunci nu numai că ai încălcat regula 60:40, ci nici nu mai e vorba de *jocul* de-a hoţii şi vardiştii, ci de faptul că îţi ţii sora captivă. Cu alte

cuvinte, acum comportamentul vostru reciproc este guvernat de FRICĂ sau de FURIE, şi nu de JOACĂ. Aceeaşi afirmaţie e valabilă şi în cazul celorlalte jocuri enumerate mai sus. „De-a doctorul şi pacientul” este un joc până ce se transformă în act sexual real, apoi e guvernat de DORINŢA SEXUALĂ. Faptul că JOACA rămâne, cum s-ar spune, între toate celelalte emoţii instinctuale – experimentându-le şi învăţându- le limitele – este probabil motivul pentru care nu se poate identifica un singur circuit cerebral pentru JOACĂ. Probabil că sunt implicate toate[31]. Totuşi, cine se îndoieşte că joaca este un instinct primar ar trebui să citească minunata carte scrisă de Sergio şi Vivien Pellis *The Playful Brain—.*

Nu ne place întotdeauna să recunoaştem că oamenii, asemenea altor mamifere, îşi revendică teritorii şi formează ierarhii sociale cu reguli clare în mod natural. (Regulile ce guvernează comportamentul primatelor sunt remarcabil de complexe.) Structura familiilor, a clanurilor, a armatelor şi chiar a ţărilor – aproape a oricărui grup social – este fără îndoială ierarhică şi teritorială, iar asta se vede în orice perioadă istorică. Pe cât statutul social al unui individ în interiorul grupului este mai înalt, pe atât mai mult acces are individul la resursele din teritoriul controlat de grup. Această observaţie nu ţine de preferinţa personală, ci e un fapt. Dacă nu recunoaştem asemenea fapte, nu ne putem ocupa de ele. Faptul că există pulsiuni emoţionale nu înseamnă că nu le putem controla – că suntem obligaţi să respectăm „legea junglei” –, dar ne asumăm un risc

dacă le ignorăm.

E uşor de observat modul în care JOACA, mai ales, creează reguli sociale. Ele reglează comportamentul de grup, protejându-ne astfel de excesele nevoilor noastre individuale. E uşor de observat şi modul în care regulile sociale încurajează forme complexe de comunicare şi, prin urmare, felul cum contribuie la apariţia gândirii simbolice. Calitatea de „ca şi cum" a circuitului JOACĂ sugerează că joaca ar putea fi chiar precursorul biologic al gândirii în general (adică a acţiunii virtuale versus acţiunea reală; vezi mai sus). Unii cercetători consideră că şi *visul* este JOACĂ nocturnă, adică testarea emoţiilor instinctuale într-o lume „ca şi cum". E interesant că, în tulburările comportamentale în somnul paradoxal, când paralizia motoare ce însoţeşte în mod normal visele dispare din cauza unor leziuni la nivelul trunchiului cerebral, pacienţii (şi animalele din cadrul unor experimente) *joacă* fizic diferitele stereotipuri instinctuale – de exemplu, fugă, încremenire, atac de prădător şi atac afectiv.

Sper că acum se înţelege că, deşi afectele corporale au o anumită intensitate şi o imediateţe ce le fac să pară într-un anumit sens de neratat, afectele emoţionale acţionează şi ele prin sentimente conştiente. Deşi nu le recunoaştem întotdeauna ca atare, ele reglează mai tot comportamentul nostru voluntar prin diversele lor imbolduri şi izbucniri interioare. Comportamentul voluntar constă, în esenţă, în a face alegeri de tip aici-şi-acum. Cum să faci alegeri fără ca ele să se bazeze pe vreun sistem de evaluare ce îţi spune care opţiune este

mai bună sau mai proastă? Cu aceste valori contribuie la comportament faptul de a simţi.

Totuşi, deoarece stările noastre emoţionale din fundal nu sunt recunoscute întotdeauna de creierul *cognitiv* şi, astfel, nu sunt întotdeauna declarate într-un mod autoreflexiv, noi nu vedem tiparul general până când nu privim în urmă şi punem lucrurile cap la cap. Aici, când spun „noi”, nu mă refer numai la nespecialişti, care au tot dreptul să nu cunoască datele experimentale. Deşi e adevărat că trecem cu vederea prea uşor rolul esenţial al sentimentelor în viaţa de zi cu zi, de fapt mă refer la opiniile general acceptate în ştiinţele cognitive contemporane. Specialiştii în ştiinţe cognitive de obicei trec cu vederea sentimentele.

Dar, după cum se va vedea, dintr-o perspectivă neurologică, în capitolul următor, orice descriere ştiinţifică a conştiinţei ce ignoră rolul fundamental al sentimentelor ratează evenimentul principal.

*1 Panksepp (1998). Nu folosesc termenul „homeostatic” în accepţiunea lui Panksepp pentru subtipul interoceptiv de afect corporal deoarece, după cum vom vedea imediat, toate afectele sunt homeostatice. Din acest motiv, utilizarea mai restrânsă a termenului la Panksepp poate fi derutantă.

6

Sursa

O ipoteză fundamentală în neuropsihologie – bazată pe metoda clinico-anatomică – este că, dacă o anumită funcţie psihică e îndeplinită de o anumită regiune cerebrală, atunci o leziune totală la nivelul ariei respective trebuie să ducă la pierderea totală a acelei funcţii. După cum am arătat, când e vorba de conştiinţă, cortexul nu trece acest test. Dar situaţia este şi mai proastă pentru teoria conform căreia conştiinţa apare exclusiv în cortex: unele leziuni *din altă parte* a creierului o anulează complet – şi încă leziuni foarte mici.

Fiziologii Giuseppe Moruzzi şi Horace Magoun au stabilit în urmă cu peste 70 de ani că conştiinţa se pierde la pisici în urma unor mici incizii care deconectează cortexul de nucleul „reticulat” (ca o plasă) al trunchiului cerebral[1]. Acest nucleu are probabil vreo 525 de milioane de ani, întrucât este comun tuturor vertebratelor – de la peşti la oameni. De la descoperirea lui de către Moruzzi şi Magoun, cercetătorii confirmă la tot felul de specii că leziuni relativ mici la nivelul acestui nucleu – cunoscut sub

denumirea tehnică de sistem reticulat activator – cauzează comă. De exemplu, David Fischer şi colegii lui au identificat recent la pacienţii umani cu accident vascular în trunchiul cerebral o regiune mică „specific comatoasă” de doi milimetri pătraţi, în partea superioară a tegmentului mezopontin (vezi figura 1)2.

Există două explicaţii posibile. Prima este că acest centru al trunchiului cerebral cu o mare densitate de noduli e locul unde apare conştiinţa: izvorul ascuns al minţii, sursa esenţei sale. Şi sunt de această părere, ca şi Jaak Panksepp. A doua explicaţie ar fi că acesta e precum cablul de alimentare pentru un televizor: necesar, dar nu suficient şi nu prea te ajută dacă vrei să înţelegi cum funcţionează televiziunea. Aceasta este opinia predominantă.

Să presupunem că a doua opţiune este adevărată. Prin stimularea trunchiului cerebral, ne-am putea aştepta să putem să pornim şi să oprim conştiinţa. În cel mai bun caz, am putea să o atenuăm în anumite feluri, aşa cum o fluctuaţie de tensiune face ecranul televizorului să pâlpâie. Nu ne-am aştepta să refacă din mers emisia curentă. Şi totuşi, un electrod implantat într-un nucleu reticulat din trunchiul cerebral al unei femei de 65 de ani (pentru tratarea bolii Parkinson) a generat următorul răspuns remarcabil:

> Chipul pacientei a exprimat o tristeţe profundă în mai puţin de cinci secunde [.]. Deşi era încă alertă, pacienta s-a lăsat spre dreapta, a început să plângă şi a comunicat verbal sentimente de tristeţe, vinovăţie, inutilitate şi deznădejde, precum „Mă prăbuşesc în

capul meu, nu mai vreau să trăiesc, nu vreau să mai văd, să mai aud, să mai simt nimic...”. Când a fost întrebată de ce plânge şi dacă o doare ceva, a răspuns: „Nu, m-am săturat de viaţă, m-am săturat. Nu mai vreau să trăiesc, mi-e silă de viaţă. Totul e inutil, mereu mă simt inutilă, lumea asta mă sperie”. Când a fost întrebată de ce e tristă, a răspuns: „Am obosit. Vreau să mă ascund într-un colţ... Îmi plâng de milă, desigur... N-am nicio speranţă, de ce vă mai deranjez?” [...]. Depresia a dispărut în mai puţin de 90 de secunde după oprirea stimulării. În următoarele cinci minute, pacienta s-a aflat într-o stare uşor hipomaniacală, a râs şi a glumit cu examinatorul, trăgându-l în glumă de cravată. Şi-a amintit întregul episod. Stimularea [într-o altă regiune cerebrală, care era adevărata ţintă a electrodului] nu a generat acelaşi răspuns psihiatric[3].

Această pacientă nu avea istoric de simptome psihiatrice de vreun fel.

Aceste afirmaţii sunt valabile şi în cazul stimulării sau blocării *chimice* a acestor nuclei din centrul trunchiului cerebral. Majoritatea antidepresivelor – activatori ai serotoninei – acţionează asupra unor neuroni ai căror corpi celulari se află într-o regiune a sistemului reticulat activator numită nucleii rafeului (vezi figura 1). Acolo e „sursa” serotoninei. Antipsihoticele – blocanţi ai dopaminei – acţionează asupra unor neuroni a căror sursă e în altă parte a sistemului reticulat activator: aria tegmentală ventrală (vezi figura 2). Afirmaţia e valabilă şi în cazul

anxioliticelor – multe dintre acestea blocând o substanţă numită noradrenalină, care este produsă de neuroni ce îşi au sursa în *locus coeruleus complex* (tot figura 1), o altă parte a sistemului reticulat activator. Toţi aceşti neuroni sunt aglutinaţi în centrul reticulat al trunchiului cerebral. Psihiatrii nu ar testa această regiune cerebrală dacă ea doar ar porni şi ar opri conştiinţa. Dacă ea ar face doar acest lucru, atunci i-ar interesa numai pe anestezişti.
Aşadar, a doua opinie trebuie să fie greşită.

Neuroimagistica funcţională a creierului în unele stări emoţionale indică aceeaşi concluzie. Tomografia cu emisie de pozitroni în timpul stărilor de MÂHNIRE, CĂUTARE, FURIE şi FRICĂ, de exemplu, arată că activitatea metabolică cea mai intensă se desfăşoară în centrul trunchiului cerebral (şi în alte zone subcorticale; vezi figura 9), iar cortexul arată *deactivare.* Imagistica prin rezonanţă magnetică funcţională în timpul orgasmului relevă acelaşi lucru: activitatea hemodinamică ce se corelează cu această stare intens afectivă se localizează aproape exclusiv în mezencefal[4].

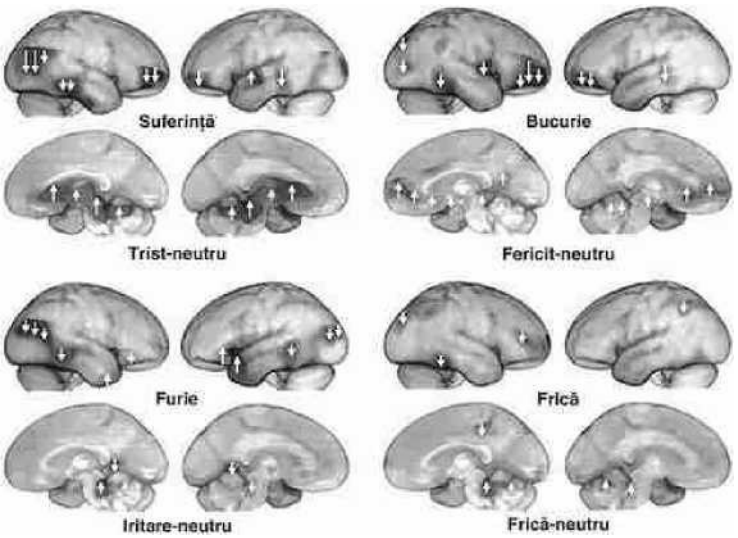


Figura 9. Imagini din tomografii cu emisie de pozitroni a patru
stări emoţionale (prin amabilitatea lui Antonio Damasio).
Săgeţile ascendente indică regiuni cu activitate *crescută,* iar
săgeţile descendente indică regiuni cu activitate *scăzută.* Zona
evidenţiată din imaginea „Bucurie” pare că arată activarea
sistemului CĂUTĂRII.

Studiile asupra unor leziuni, stimularea cerebrală profundă, manipularea farmacologică şi neuroimagistica funcţională indică aceeaşi concluzie: centrul reticulat al trunchiului cerebral generează *afecte.* Prin urmare, aparent, singura parte a creierului despre care ştim că este necesară pentru activarea conştiinţei în ansamblul său are o influenţă la fel de puternică asupra altei funcţii psihice, şi anume sentimentele. În capitolul anterior, am arătat că sentimentele răzbat în întreaga

experienţă conştientă: indiferent ce altceva e menită să facă, una dintre principalele sarcini ale conştiinţei pare să constea în a avea şi a gestiona sentimente (care vin din *interiorul* nostru şi ne reglează nevoile biologice). Dar acum pare că sursele neurologice ale afectului şi ale conştiinţei cel puţin se întrepătrund în mare măsură şi, de fapt, s-ar putea să coincidă. Contrar opiniei empiriste clasice, conform căreia conştiinţa ne parvine prin simţuri, şi contrar afirmaţiei citate din Meynert ce se bazează pe această opinie, se pare că creierul *chiar* „radiază propria căldură”.

Cum ar trebui să numim mediul de bază, această misterioasă materie a minţii ce pare să izvorască în interiorul nostru? Nu o putem numi „stare de veghe”, aşa cum o numea Zeman, întrucât asta ne-ar impune să descriem visul drept un tip de trezie, ceea ce este absurd. Nu o putem numi nici „nivel” cantitativ, ca Moruzzi şi Magoun, întrucât faptele descrise mai sus arată că ea are trăsături intens calitative.

Aşadar, să încercăm al treilea termen folosit în literatura de specialitate: „activare”. Mi se pare un cuvânt bun, neutru. Atât starea de veghe, cât şi visul implică activare; şi nu exclude calitatea – ca în cazul termenului „nivel”. De fapt, „activare” sugerează categoric faptul de a simţi.

Dar ce *este* activarea? Am vorbit despre ea în termeni făţiş behaviorişti – de exemplu, deosebirea dintre comă, starea vegetativă şi starea de veghe cu reacţii complete. De obicei, e măsurată cu ajutorul scalei de comă Glasgow: testarea răspunsului ocular al pacientului, a răspunsului verbal şi a răspunsului motor

la instrucţiuni şi la durere. Totuşi, ea poate fi definită şi din punct de vedere psihologic.

O electroencefalogramă (EEG) produce reprezentări grafice ale activităţii electrice corticale. Când e lăsat în voia lui (adică este deconectat de la sistemul reticulat activator, chiar şi atunci când procesează inputuri senzoriale), cortexul produce tiparul undelor *delta,* o serie de unde cu amplitudine mare ce apar de aproximativ două ori pe secundă (adică o frecvenţă de 2 Hz). Când este stimulat de sistemul reticulat activator în absenţa unui input senzorial, cortexul produce de regulă ritmul *teta* (4-7 Hz)[5] sau ritmul *alfa* (unde desincronizate la frecvenţe de 8-13 Hz; „desincronizat" înseamnă neregulat). Când prelucrează în mod activ informaţii din exterior, cortexul are de obicei ritm *beta* (unde desincronizate, de amplitudine foarte mică, cu frecvenţe de 14-24 Hz) sau *gama* (unde de mică amplitudine, cu frecvenţe foarte mari, de 25-100 Hz). Gama este ritmul asociat de regulă cu conştiinţa.

Astăzi, activarea fiziologică poate fi măsurată şi cu ajutorul neuroimagisticii funcţionale, ce ilustrează la propriu activitatea cerebrală prin topografierea unor ritmuri regionale de modificare a ratelor metabolice. Figura 3 ilustrează această tehnică, cu referire la diferite faze ale somnului. Rândul de jos din figură arată activarea somnului REM, care este de obicei (deşi, după cum am arătat, nu exclusiv) asociat cu conştiinţa în timpul visului şi care îşi are sursa în trunchiul cerebral superior. Neuroimagistica unora dintre stările emoţionale de bază ale conştiinţei descrise mai sus (vezi figura 9) şi ale orgasmului arată acelaşi lucru: activarea

trunchiului cerebral.

Cortexul devine conştient numai în măsura în care este activat de trunchiul cerebral. Relaţia dintre cele două este ierarhică; conştiinţa corticală *depinde* de activarea din trunchiul cerebral. Acesta este motivul pentru care ritmul undelor delta prezentat primul de sus în figura 10 – care nu este asociat cu comportamentul conştient – este generat de activitatea corticală intrinsecă şi tot din acest motiv ritmul gama prezentat ultimul în figura 10 – care este puternic asociat cu conştiinţa – poate fi impulsionat exclusiv de sistemul reticulat activator[6]. Şi tot din acest motiv scăderea activării fiziologice ilustrată pe rândul de sus din figura 3 coincide cu atenuarea conştiinţei şi instalarea somnului, iar activarea intensificată ilustrată pe rândul de jos coincide cu reapariţia conştiinţei în timpul visului. Acestea nu sunt fapte controversate.

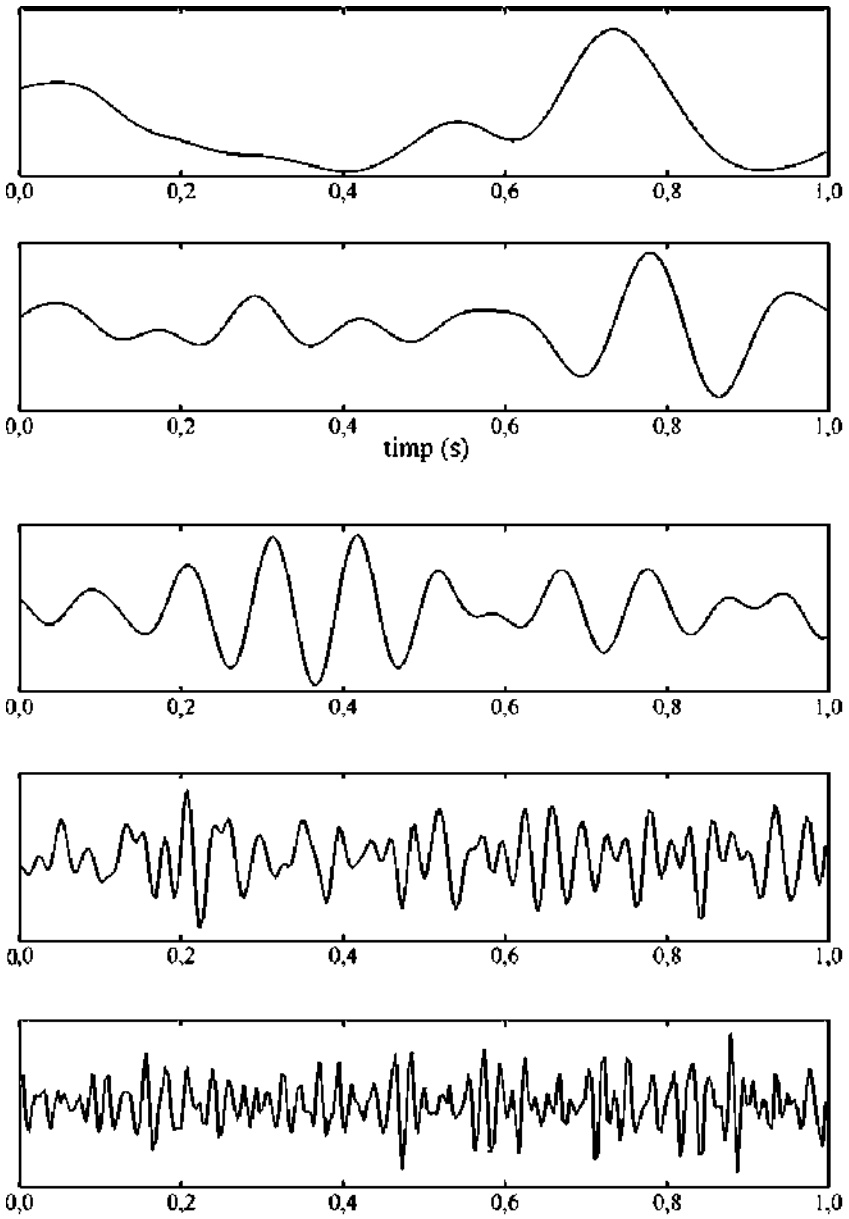


Figura 10. EEG-uri cu ritmuri tipice ale activităţii corticale.
Ritmul cel mai puţin activat (delta) este primul de sus, iar
ritmul cel mai activat (gama) ultimul. Ritmurile intermediare,
de sus în jos, sunt teta, alfa şi beta.

Să examinăm mai atent mecanismele cerebrale

propriu-zise implicate. Pentru a face acest lucru, trebuie să vorbesc despre o deosebire fundamentală între două moduri în care neuronii comunică între ei. Se dovedeşte că această deosebire este importantă pentru conştiinţă.

Majoritatea celor oarecum interesaţi de creier ştiu că neuronii transmit mesaje de-a lungul unor reţele complexe. Acest proces se numeşte *transmisie sinaptică,* deoarece implică transmiterea unor mesaje prin intermediul sinapselor, structurile prin care neuronii îşi transmit semnale (cuvântul „sinapsă” vine din termenul grecesc pentru „a prinde împreună”). Transmisia sinaptică foloseşte molecule numite *neurotransmiţători,* care sunt trecute de la un neuron la următorul fie excitând neuronul postsinaptic, fie inhibându-l, în funcţie de molecula în cauză (glutamatul şi aspartatul sunt neurotransmiţători excitatori, iar acidul gamaaminobutiric [GABA] este inhibitor). Dacă neuronul din aval este excitat de un val de neurotransmiţători, îşi transmite propriile molecule următorilor neuroni din reţea. În caz contrar, nu transmite. Apoi, moleculele transmise se degradează rapid sau sunt luate înapoi în neuronul presinaptic, pentru a se limita durata efectului lor – un proces numit „recaptare”.

Transmisia sinaptică este ţintită, binară (da/nu) şi rapidă. Acest aspect al funcţionării creierului aminteşte cel mai mult de calculul digital, astfel explicându-se poate de ce suscită atât de mult interesul specialiştilor în neuroştiinţe care au înclinaţii spre calculatoare. Ea apare în tot sistemul nervos, inclusiv în cortex. Dar nu

este intrinsec conştientă. Cu alte cuvinte, acest tip de neurotransmisie apare în cortex, indiferent dacă este conştientă sau nu. Şi nu are aproape nimic de-a face cu activarea.

Puţini ştiu însă că transmisia sinaptică are loc sub influenţa constantă a unui proces fiziologic complet diferit. Acest alt tip de activitate neuronală se numeşte *modulare postsinaptică.* Spre deosebire de transmisia sinaptică, modularea postsinaptică este dezordonată, chimică şi foarte diferită de ce se întâmplă într-un computer obişnuit. Ea apare în mod endogen din sistemul reticulat activator (şi din alte structuri subcorticale şi chiar din unele structuri nonneurologice ale corpului) şi are legătură cu activarea.

Jucătorii principali în acest proces sunt o clasă de molecule numite *neuromodulatori.* Spre deosebire de neurotransmiţători, aceste molecule sunt răspândite în tot creierul – adică sunt eliberate în apropierea unor populaţii întregi de neuroni, şi nu spre sinapse individuale*[1]. În loc să transmită mesaje de-a lungul unor „canale” specifice, ele inundă legături din reţea, reglând astfel „starea” generală a cortexului. De exemplu, cortexul se află în stări diferite pe rândurile de sus şi de jos ale imaginilor din figura 3 (somn cu unde lente versus somn REM) şi în cele patru stări emoţionale ilustrate în figura 9 (MÂHNIERE versus CĂUTARE versus FURIE versus FRICĂ). În fiecare dintre aceste stări, el prelucrează diferit informaţiile. Astfel, dacă cineva te strigă pe nume când dormi, reacţionezi foarte diferit faţă de atunci când eşti complet treaz[7]. De asemenea, gândeşte-te ce reacţie ai

faţă de un străin care se apropie dacă eşti într-o stare de CĂUTARE versus FRICĂ: în prima stare poate că l-ai saluta pe străin şi chiar ai începe o conversaţie cu el, iar în a doua stare probabil ai privi în altă parte, sperând că nu te observă.

Această deosebire dintre „canal” şi „stare” este o prescurtare utilă pentru cele două moduri în care neuronii comunică între ei[8]. *Starea* cortexului afectează intensităţile diferenţiale ale transmiterii mesajelor ce continuă în interiorul *canalelor* sale; într-un anumit sens, starea reglează cât de „sonor” comunică diferitele canale (vezi figura 11). Acesta este motivul pentru care acelaşi sunet (de exemplu, cineva care te strigă pe nume) este generalizat în cortex în timpul stării de veghe, dar este izolat în cortexul auditiv în timpul somnului, iar prezenţa unui străin activează o reţea cerebrală în timpul stării de FRICĂ şi alta în timpul CĂUTĂRII.

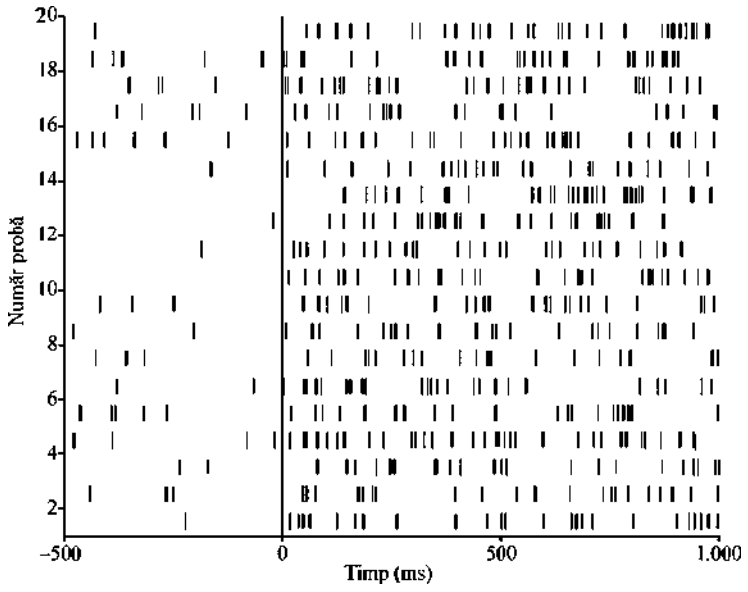


Figura 11. Imaginea de mai sus este o reprezentare

grafică a
măsurătorilor potenţialelor de acţiune *(spike trains)* a 20 de
neuroni (reprezentaţi pe axa verticală *y)* într-un interval de 1,5
secunde (reprezentat pe axa orizontală *x*) în timpul expunerii la
un stimul vizual. Expunerea la stimul intervine la momentul 0
pe axa *x* (reprezentată de a doua linie verticală). În acest punct,
neuronii, care au o frecvenţă de descărcare iniţială de 6 Hz (în
medie, când nu există stimul), îşi intensifică frecvenţa de
descărcare medie la 30 Hz. Un *spike train* este o succesiune
temporală în care un neuron se descarcă (= acţiune) şi nu se
descarcă (= latenţă). Acest lucru poate fi reprezentat ca o

secvenţă digitală de informaţii: „1” pentru o acţiune şi „0”
pentru o latenţă. De exemplu, un *spike train* ar putea fi
reprezentat în cod ca „001111101101”. Aici, primele două
zerouri reprezintă intervalul de latenţă dintre expunerea la
stimul şi prima acţiune. E important de menţionat că
frecvenţele de descărcare nu sunt determinate numai

de
stimul, ci ele apar dintr-o interacţiune între stimul, potenţarea
sau inhibarea pe termen lung a neuronilor (adică dacă stimulul
este familiar sau nefamiliar) şi nivelul curent al modulării lor.
Ajustarea modulării postsinaptice este efectul activării asupra
neuronilor. Din acest motiv, stimulul ilustrat aici ar putea să nu
evoce niciun răspuns din partea aceloraşi neuroni atunci când
activarea este modulată descendent.

Aceasta este esenţa a ceea ce numim activare. De observat însă că activarea corticală poate fi modulată atât ascendent, *cât şi descendent,* până la suprimarea totală a transmisiei, după cum se întâmplă în fiecare noapte când ne culcăm (motiv pentru care unii fiziologi preferă termenul „modulare” în loc de „activare”). Prin urmare, activarea determină ce impulsuri sinaptice vor fi transmise şi cât de puternic, ca în exemplul în care eşti strigat pe nume când dormi versus când eşti treaz.

Transmisia sinaptică este binară (pornit/oprit, da/nu, 1/0), dar neuromodularea postsinaptică *ierarhizează* probabilitatea ca un anumit set de neuroni să se descarce. Modifică şansele statistice ca ceva să se întâmple în interiorul lor. Această ajustare probabilistă şi analoagă a frecvenţelor de descărcare se face prin receptori aflaţi în mai multe locuri de-a lungul unui

neuron. Spre deosebire de neurotransmiţători, neuromodulatorii au efecte relativ lente şi de durată – nu numai pentru că substanţele se menţin mai mult timp, ci şi pentru că, în cazul canalelor care se descarcă mai frecvent, creşte probabilitatea de a se descărca în viitor. Dacă activezi o anumită parte a reţelei, ea va rămâne oarecum activată până când este modulată descendent. Acest lucru influenţează neuroplasticitatea şi constituie o parte importantă a modului în care funcţionează învăţarea. Stările de activare înregistrează diversele lecţii pe care le învăţăm mai profund în canalele creierului. Astfel, de exemplu, probabilitatea să-ţi aminteşti de o călătorie este mai mare dacă încerci cu nerăbdare să găseşti o destinaţie nefamiliară decât dacă tot călătoreşti în acelaşi loc din obişnuinţă, pe pilot automat.

De unde provin neuromodulatorii? Din tot corpul, inclusiv din glandele hipofiză, suprarenală, tiroidă şi sexuale (care produc diverşi hormoni) şi din hipotalamus (care produce numeroase peptide). Dar sursa principală a „activării” din punctul de vedere al creierului este sistemul reticulat activator. Activarea reticulată a trunchiului cerebral eliberează cei mai cunoscuţi cinci neuromodulatori: *dopamina* (având sursa principală în aria tegmentală ventrală şi în substanţa neagră), *noradrenalina* (având sursa principală în *locus coeruleus complex), acetilcolina* (având sursa principală în nucleii tegmentului mezopontin şi ai prozencefalului bazal), *serotonina* (având sursa principală în nucleii rafeului) şi *histamina* (având sursa principală în corpii mamilari ai

hipotalamusului). Nu întâmplător unele dintre aceste substanţe au reprezentat subiecte importante de discuţie în capitolele anterioare. Fiecare dintre molecule şi receptorii lor aferenţi, care au multe subtipuri, este responsabil pentru un aspect diferit al activării modulatoare. Pe lângă acestea cinci, există numeroase altele – în principal hormoni şi peptide cu acţiune lentă (în creier sunt peste o sută) – care modulează sisteme neuronale cu grad mare de specificitate.

Efectul tuturor acestor modulatori este determinat de prezenţa sau absenţa receptorilor relevanţi şi a subtipurilor de receptori. Cu alte cuvinte, deşi neuromodulatorii sunt răspândiţi în întregul creier, ei influenţează numai celulele care au receptorii relevanţi. Nu trebuie să ne gândim la activare ca la un proces brut sau nediscriminatoriu. Ea are foarte multe aspecte şi se desfăşoară sub multe dimensiuni, atât spaţial, cât şi temporal. Mesajele potenţiale purtate de neuromodulatori circulă prin numeroasele reţele cerebrale, dar (asemenea analgezicelor) sunt utilizate numai acolo unde este nevoie de ele. De exemplu: în timp ce citeşti acest text, tot felul de stimuli de fond sunt înregistraţi de sistemele tale senzoriale fără să le acorzi atenţie, dar, dacă eşti mamă şi nou-născutul tău începe să plângă, acest lucru devine prioritar faţă de atenţia pe care o acorzi cărţii şi imediat ţi-o redirecţionezi către bebeluş. Asta se datorează nivelurilor crescute ale peptidelor şi hormonilor specifici – estrogen, progesteron, prolactină şi oxitocină – care îţi circulă prin creier când ai un

bebeluş, schimbându-ţi starea (adică activează sistemul GRIJĂ al creierului).

Neuromodulatorii pot să moduleze (ascendent sau descendent) numai semnale care sunt prezente – cu alte cuvinte, canale active în momentul respectiv. Ei sunt eliberaţi în mod difuz, dar influenţează numai neuronii care (1) au receptorii relevanţi şi (2) sunt activi în momentul respectiv. Acelaşi modulator are efecte diferite asupra unor clase diferite de receptori, în locuri diverse. Se obţine mai multă specificitate de către modulatorii eliberaţi nu numai de trunchiul cerebral (şi de sistemul endocrin), ci şi în circuitele neuronale. Cei mai cunoscuţi cinci modulatori ai activării pe care i-am enumerat mai sus îşi au sursa principală în trunchiul cerebral reticulat. Hormonii îşi au sursa în alte părţi ale corpului şi ajung la creier în principal prin fluxul sangvin. Modulatorii mai specifici sunt peptidele, multe dintre ele avându-şi sursa în hipotalamus. Fiecare peptidă face ceva diferit, din nou în funcţie de tipul şi localizarea receptorului. Multe dintre ele, care modulează emoţiile de bază în moduri foarte precise, sunt enumerate în notele de final ale capitolului 5.

Până acum în acest capitol m-am ocupat de două întrebări: *de unde* vine activarea din punct de vedere anatomic şi *cum* apare ea din punct de vedere fiziologic. Răspunsurile sunt că activarea este generată în principal, dar nu exclusiv în trunchiul cerebral şi în hipotalamus şi activează prozencefalul modulând neurotransmisia. Pentru a reaminti din capitolul 5, *motivul pentru care* apare activarea este că ea răspunde

unor solicitări endogene de travaliu asupra psihicului. Aceste solicitări iau forma multor semnale „de eroare” ce converg în centrul creierului. În majoritatea cazurilor, asemenea semnale sunt gestionate în mod automat şi inconştient, solicitându-se răspunsuri conştiente numai în situaţii în care răspunsurile automate nu sunt suficiente. Iar astfel ajung la întrebarea principală a capitolului de faţă: unde apare această trecere aparent magică de la reflexul automat la sentimentul volitiv?

Ceea ce urmează să descriu şi să explic nu este definitiv, din cauză că faptele încă nu sunt cu totul clare. Cunoaşterea actuală din neuroştiinţe este însă suficient de avansată pentru a face o descriere în linii mari a imaginii de ansamblu. Pionierul acestor cercetări a fost Jaak Panksepp, urmat de Antonio Damasio. Damasio are darul de a vedea imaginea de ansablu, dar specialistul în neuroştiinţe căruia consider că îi suntem cei mai îndatoraţi aici, cu privire la modul în care creierul face saltul misterios de la reflexul automat la acţiunea voluntară, este Björn Merker. El studiază activarea trunchiului cerebral şi mecanismele tropismului la o mare diversitate de specii de vertebrate, inclusiv păsări, rozătoare, feline şi primate.

Aparent, trecerea de la starea de veghe vegetativă la activarea afectivă depinde de integritatea unui conglomerat mic, foarte compact de neuroni din jurul canalului central al mezencefalului – *substanţa cenuşie periapeductală* (PAG), unde converg *toate* circuitele cerebrale afective (vezi figura 6 pentru localizarea acestei structuri). Prin urmare, pacienţii cu leziuni

localizate în PAG „au privirea pierdută într-un spaţiu psihoafectiv gol”[9]:

> Leziunile extinse la nivelul PAG [cauzează] o deteriorare spectaculoasă a tuturor activităţilor conştiente [...]. De exemplu, primele studii în care au fost inseraţi electrozi pentru producerea unor leziuni din al patrulea ventricul ascendent prin apeduct spre limita caudală a diencefalului au evidenţiat deficienţe impresionante ale conştiinţei la pisici şi maimuţe, după cum au arătat *eşecul lor de a manifesta orice comportament intenţional aparent* şi *lipsa generală de răspuns la stimuli emoţionali*[10]*.* Deşi unele forme de leziuni în oricare alte zone cerebrale superioare pot să afecteze „instrumentele [cognitive] ale conştiinţei”, de regulă ele nu deteriorează baza intenţionalităţii în sine. Leziunile PAG reuşesc acest lucru cu cea mai mică
> distrugere absolută a ţesutului cerebral[11].

Starea descrisă este „starea vegetativă”. Principala diferenţă dintre aceasta şi comă este menţinerea stării de veghe. Dar ritmul circadian somn/veghe este doar o altă funcţie autonomă[12]. Din acest motiv, starea vegetativă se numeşte şi „stare de veghe *fără răspuns la stimuli”* – un aparent oximoron ce relevă deosebirea esenţială dintre starea de veghe (vegetativă) şi activare (afectivă) –, ceea ce Panksepp numeşte „intenţionalitate” în citatul de mai sus[13]. Acesta este un alt motiv pentru care eu prefer termenul „activare” în loc de „stare de veghe” sau de „nivel” al conştiinţei.

„Activarea” poate include (şi chiar sugerează) capacitatea de răspuns emoţional şi intenţionalitatea, care, după cum vedem din nou aici, se află în centrul comportamentului conştient[14]. Activarea afectivă permite *voliţiunea.* Iată ce adaugă PAG la funcţionarea automată, vegetativă.

Cum face acest lucru? PAG nu este parte din sistemul reticulat activator, deşi se învecinează şi e dens interconectată cu el[15]. Principala diferenţă dintre aceşti nuclei şi PAG este direcţia fluxului de informaţii dintre ele şi prozencefal. Sistemul reticulat activator îşi exercită influenţa în principal *ascendent,* spre cortex, iar cortexul doar transmite semnale *descendent,* spre PAG*[2].

PAG este punctul final în care se adună toate circuitele cerebrale afective. Aşadar, prozencefalul este activat de sistemul reticulat activator, iar PAG este activat (ca să zic aşa) de prozencefal. Ne-am putea gândi la sistemul reticulat activator şi la PAG ca la originea şi, respectiv, destinaţia activării prozencefalului.

În consecinţă, PAG este conceptualizată drept punctul terminus al reţelei „descendente” pentru afecte, pentru a fi pusă în contrast cu reţelele afective „ascendente” şi „modulatoare” ale nucleilor care monitorizează corpul şi ale sistemului reticulat activator din trunchiul cerebral[16]. Asta înseamnă că PAG este principalul centru de *output* al tuturor circuitelor afective, canalizând informaţiile spre efectorii musculoscheletali şi viscerali care generează „emoţie

propriu- zisă”—. (Aici citez direct din analize de specialitate, ca să mă asigur că ilustrez cu acurateţe modul în care înţelegem aceste funcţii.) Reţeaua descendentă pentru afect este „implicată în acţiuni motoare specifice invocate în emoţii, precum şi în controlul frecvenţei cardiace, respiraţiei, vocalizării şi comportamentului de împerechere”[18]. Rolul PAG în această reţea este de *„interfaţă* asupra stimulilor proeminenţi între prozencefal şi trunchiul cerebral inferior”[19]. În această privinţă, PAG este conceptualizată drept centru pentru *„echilibrarea sau transmiterea* informaţiilor legate de predominanţa supravieţuirii”[20]. Altfel spus, PAG funcţionează *„orchestrând* diferite strategii de *coping* când este expusă la stresori externi”[21]. PAG „constituie un *punct de întâlnire* masiv al sistemelor neuronale care generează emotivitate”[22]. Astfel, ea joacă rolul central în „apărarea homeostatică a răspunsului individului, *integrând* informaţiei aferente de la periferie şi informaţii din centre superioare”[23].

Încerc să reformulez cât mai direct: *toate circuitele afective converg în PAG, care este principalul centru de output pentru sentimente şi comportamente emoţionale.* Din acest motiv, „stimularea electrică la intensităţi mai joase a acestei zone cerebrale va activa la animale o mai mare diversitate de acţiuni emoţionale coordonate decât stimularea oricărei alte regiuni cerebrale”[24].

Ca aspect, coloanele slab diferenţiate ce constituie PAG înconjoară canalul central al trunchiului cerebral pe o lungime de 14 mm. Canalul central, prin care circulă un lichid incolor (lichidul cefalorahidian), este

apeductul lui Sylvius – numit după anatomistul din secolul al XVII-lea care l-a descris pentru prima dată. Localizarea sa în centrul mezencefalului îi dă numele PAG: „substanţa cenuşie periapeductală” înseamnă, pur şi simplu, „substanţa cenuşie care înconjoară canalul”[25]. Acest centru primitiv al creierului este cu adevărat ceea ce a descris Freud drept „straturile sale intime”. Se împarte în două grupuri de coloane funcţionale[26]. Una dintre ele, cea din spate, este pentru „strategii de *coping” active* sau pentru comportamente defensive precum reacţiile de tip luptă-sau-fugă, tensiunea arterială crescută şi reducerea durerii pe căi ce nu au legătură cu opioidele[27]. Aici se termină circuitele FRICII, FURIEI şi PANICII/MÂHNIRII. Coloana din faţă este pentru strategii defensive/de *coping pasive* precum încremenirea cu hiporeactivitate, „comportamentul de bolnav” pe termen lung, tensiunea arterială scăzută şi ameliorarea durerii cu opioide[28]. Circuitele DORINŢEI SEXUALE, GRIJII şi CĂUTĂRII se termină în această coloană din faţă.

PAG este *calea comună finală* spre outputul afectiv. Prin urmare, ea trebuie, într-un cuvânt, „să aleagă” ce e de făcut în continuare, odată ce diferitele circuite afective şi comportamentele condiţionate aferente şi-au adus contribuţiile la acţiune[29]. Ea trebuie să facă aceste alegeri prin evaluarea semnalelor de eroare reziduale ce îi sunt retransmise prin sistemele afective. Trebuie să le analizeze cererile concurente în funcţie de imperativele biologice supreme ale supravieţuirii şi reproducerii, fiecare semnal de eroare comunicându-i nevoia componentă. Pe scurt, PAG trebuie *să*

stabilească priorităţile pentru următoarea secvenţă de acţiune.

Dar acţiunile prioritare nu pot fi determinate exclusiv în funcţie de nevoi. Nevoile trebuie să fie *contextualizate* nu numai reciproc, ci şi de oportunităţile predominante.

Să luăm exemplul controlului respirator, descris mai sus. Să ne gândim mai întâi la contextul intern: dacă simt alarma de sufocare şi sete în acelaşi timp, alarma de sufocare trebuie să fie prioritizată în faţa setei, dar setea ar putea să aibă prioritate, de pildă, în faţa tristeţii. În al doilea rând, să ne gândim la contextul extern: dacă resimt o alarmă de sufocare, contextul îmi va spune dacă am nevoie – de exemplu – să elimin o obstrucţie a căilor respiratorii sau să ies dintr-o încăpere plină cu dioxid de carbon. Contextul extern în care activez rămâne relevant pe măsură ce evenimentele se desfăşoară, iar eu tot iau hotărâri privind ce să fac în continuare.

Să ne amintim definiţia afectului din capitolul 5; ea include multe lucruri pe care oamenii nu le consideră de regulă a fi afective. Astfel, de exemplu, alarma de sufocare şi setea nu sunt doar senzaţii corporale, ci, la fel ca anxietatea de separare, de pildă, transmit şi valori intrinsece – ceva bun şi ceva rău din punct de vedere biologic. Durerea e un exemplu util pentru a face distincţie între valenţa intrinsecă a afectului şi contextul său exteroceptiv. *Senzaţia* neplăcută de durere e ceea ce o face afectivă, iar senzaţia somatică exteroceptivă transmite *localizarea* unui stimul dureros: „durere ce vine din mâna stângă”. Aceasta nu este o distincţie

filosofică – aspectele duale pot fi manipulate separat, de exemplu, prin stimularea PAG şi, respectiv, a cortexului parietal. Prin urmare, în anumite condiţii clinice e posibil *să percepi* că te-ai înţepat cu o piuneză în mâna stângă fără *să simţi* durere[30].

Pe acest fond, Björn Merker face o observaţie profundă[31]. Deşi din punct de vedere anatomic PAG este localizată sub cortex, ea e cea mai importantă din punct de vedere funcţional. *După ce* cortexul şi alte structuri ale prozencefalului şi-au făcut treaba cognitivă – după ce şi-au adus contribuţia la acţiune –, *decizia finală* despre „ce să fac în continuare” se ia la nivelul mezencefalului. Asemenea decizii iau forma trăirilor afective, generate de PAG, care pot să aibă precedenţă asupra oricăror strategii cognitive formulate în timpul secvenţei anterioare de acţiune. În orice moment, PAG alege afectul care va determina şi va modula secvenţa următoare. De exemplu, în prima mea zi de lucru ca neuropsiholog, am decis *din punct de vedere cognitiv* să revin la paturile pacienţilor ca să le citesc fişele, dar ceea ce s-a întâmplat de fapt *din punct de vedere afectiv* este că am leşinat.

Desigur, viaţa nu constă din urgenţe nesfârşite; cel mai adesea, va prevala activitatea implicită a CĂUTĂRII (interesul şi implicarea de fond). Totuşi, în situaţii de urgenţă, numeroasele semnale de eroare ce transmit nevoi şi ajung în mezencefal nu pot fi percepute toate odată, deoarece nu poţi *să faci* tot odată. Aşadar, aceasta este esenţa funcţiei de prioritizare a PAG: care dintre aceste semnale de eroare sunt predominante, date fiind circumstanţele de moment? Care dintre

problemele mele poate fi amânată (sau gestionată automat) şi în cazul căreia trebuie să caut soluţia în funcţie de ceea ce simt?

Cerinţa că trebuie să ţii cont într-un mod firesc de circumstanţele curente presupune un aparat suplimentar. PAG îşi dă verdictul cu ajutorul unei structuri adiacente din mezencefal, numită *coliculii superiori* (vezi figura 6). Această structură se află chiar în spatele ei şi este împărţită în mai multe straturi, fiecare oferind reprezentări simple derivate din simţurile corpului. Straturile mai profunde furnizează hărţi motoare ale corpului, iar straturile superficiale sunt responsabile pentru cele spaţio-senzoriale. Împreună, ele alcătuiesc o reprezentare extrem de comprimată şi de integrată a lumii exteroceptive, provenind parţial din cortex, dar şi din regiuni subcorticale senzorimotoare precum nervul optic (din nou, vezi figura 6). Astfel, coliculii superiori reprezintă într-o formă condensată starea corpului *obiectiv* (senzorial şi motor) moment cu moment, cam aşa cum PAG îi monitorizează starea *subiectivă* (nevoia). Merker numeşte această interfaţă afectivă/senzorială/motoare dintre PAG, coliculii superiori şi zona locomotoare din mezencefal „triunghiul de decizie” al creierului[32]. Panksepp a numit-o SINELE primar, însăşi sursa fiinţei noastre capabile de simţire[33].

Prin urmare, deciziile mezencefalului despre ce să faci în continuare se bazează pe feedback din partea circuitelor afective cerebrale *alături de* hărţile senzorimotoare, fiecare dintre acestea actualizându-ne diferitele aspecte ale „felului cum stau lucrurile acum”.

Să ne amintim de exemplul în care am conştientizat brusc nevoia de a urina la sfârşitul unei prelegeri de două ore. În acel moment, simţind ocazia, nevoia mea corporală a fost percepută şi a devenit o *pulsiune* voliţională. Pe scurt, deoarece triunghiul de decizie din mezencefal ia în considerare atât condiţiile interne, cât şi pe cele externe, el prioritizează opţiunile comportamentale nu numai pe baza nevoilor curente, ci şi a oportunităţilor curente.

Stratul cel mai profund al coliculilor superiori constă într-o hartă ce controlează mişcările ochilor. Această hartă este în sine mai stabilă decât hărţile senzoriale superioare, întrucât celelalte hărţi sunt calibrate în funcţie de ea, stabilind astfel „punctul de vedere” unificat ce caracterizează experienţa perceptivă subiectivă. Acesta este motivul pentru care ne percepem pe noi înşine în interiorul unei scene vizuale stabile indiferent cât de mult ni se mişcă ochii, căci se mişcă de aproximativ trei ori pe secundă. De asemenea, scena stabilizată sugerează că ceea ce percepem asta şi este – o *scenă* –, o perspectivă construită pornind de la realitate, nu realitatea însăşi. Tot din acest motiv ne percepem pe noi înşine ca trăind în mintea noastră[34]. Voi explica mai detaliat natura virtuală a percepţiei în capitolul 10.

După cum am văzut în cazul vederii oarbe, „harta bidimensională asemenea unui ecran” a lumii senzorimotoare, cum o numeşte Merker, este inconştientă la oameni[35]. Ea nu prea conţine decât o reprezentare a direcţiei „deviaţiei-ţintă” – ţinta fiind punctul asupra căruia se concentrează fiecare ciclu de

acţiune –, producând privirea intenţionată, atenţia şi orientarea acţiunii. Brian White o numeşte hartă „a predominanţei” sau „a priorităţii”. Panksepp arată că acesta este modul în care „deviaţiile noastre de la o stare de repaus ajung să fie reprezentate ca stări de disponibilitate pentru acţiune”[36]. Mai bine de atât nu cred că se poate formula.

Conştientizarea perceptivă a lumii din jurul nostru devine posibilă cu ajutorul cortexului activat corespunzător, care (spre deosebire de conştiinţa afectivă) este ceea ce le lipseşte copiilor cu hidroanencefalie şi animalelor ce au suferit decorticare cerebrală. Coliculii superiori furnizează mapări condensate de tip aici-şi-acum ale ţintelor şi acţiunilor potenţiale, dar cortexul furnizează „reprezentări” detaliate pe care le folosim pentru a ghida fiecare secvenţă de acţiune pe măsură ce se desfăşoară. Pe lângă aceste imagini foarte diferenţiate, în prozencefalul subcortical există multe programe de acţiune *inconştiente* ce sunt numite „proceduri” şi „răspunsuri” – nu imagini. (Să ne gândim, de exemplu, la tipurile automatizate de memorie pe care ne bazăm ca să mergem pe bicicletă sau ca să parcurgem traseul spre un loc familiar.) Acestea sunt codificate în principal în ganglionii bazali *subcorticali,* amigdală şi cerebel. Amintirile nu sunt simple consemnări ale trecutului. Din punct de vedere biologic, ele sunt *despre* trecut, dar *pentru* viitor. În esenţă, toate sunt nişte *predicţii* menite să vină în întâmpinarea nevoilor noastre. Voi relua această idee importantă în capitolele următoare.

Tendinţele motoare activate prin selecţia mezencefal-afect deblochează reflexe şi instincte simple – şi asta e tot ce fac în cazul bebeluşilor, al copiilor cu hidroanencefalie şi al multor animale. Dar, după cum ştim de-acum, asemenea comportamente automate intră sub un control individualizat pe parcursul dezvoltării, prin învăţarea din experienţă. Drept urmare, răspunsurile stereotipe sunt completate de un repertoriu mai flexibil de opţiuni. Secvenţa comportamentală ce apare cu fiecare ciclu de acţiune nou se desfăşoară ascendent deasupra acestor niveluri ale controlului din prozencefal ce se extind progresiv, de la „răspunsuri" procedurale la „imagini mnezice" reprezentaţionale. Aşa apare ceea ce Merker numeşte „o lume tridimensională, panoramică, complet articulată, compusă din obiecte solide cu formă definită: lumea experienţei noastre fenomenologice familiare"[37].

Să ne amintim că memoria pe termen lung serveşte viitorului. Odată ce triunghiul de decizie din mezencefal a evaluat feedbackul comprimat ce vine din fiecare acţiune anterioară, el activează un proces de *feedforward* extins ce se desfăşoară în sens invers, prin sistemele de memorie din prozencefal, generând un *context aşteptat* pentru secvenţa motoare selectată. Acesta este produsul întregii noastre învăţări. Cu alte cuvinte, când o nevoie ne scoate în lume, *nu redescoperim lumea de fiecare dată, cu fiecare ciclu nou.* Ea activează un set de predicţii despre consecinţele senzoriale probabile ale acţiunilor noastre, în baza experienţei noastre din trecut de satisfacere a nevoii selectate în împrejurările predominante.

Apoi, acţiunea voluntară declanşează un proces de testare a *aşteptărilor* noastre în comparaţie cu consecinţele *reale* ale acţiunilor noastre. Această comparaţie generează un semnal de eroare, pe care îl folosim ca să ne reevaluăm aşteptările pe parcurs şi ca să ne ajustăm planurile de acţiune în mod corespunzător. La asta se referă comportamentul „voluntar”: să decizi ce să faci în condiţii de incertitudine, în funcţie de consecinţele fiecărei acţiuni. Asta presupune consecinţe resimţite atât afectiv, cât şi perceptiv, motiv pentru care şi semnalele de eroare afective, şi cele reziduale senzorimotoare converg în triunghiul de decizie din mezencefal. Îmi place cum numeşte Jakob Hohwy procesul psihic ce controlează comportamentul voluntar: „prezicerea prezentului”[38].

Ca să-ţi faci o idee despre cum arată din punct de vedere clinic ceea ce spun, gândeşte-te la ce s-a întâmplat cu domnul S., inginerul electronist cu psihoză Korsakoff al cărui caz l-am descris în capitolul 2. El se aştepta să vadă Johannesburgul când s-a uitat pe fereastra din cabinetul meu, lucru cauzat de tulburarea lui de memorie: nu îşi actualizase modelul predictiv astfel încât să corespundă cu evenimentele recente. Când peisajul londonez cu zăpadă pe care l-a văzut în realitate nu s-a potrivit cu aşteptările lui, domnul S. nu şi-a schimbat abordarea. A ignorat semnalul de eroare pe care îl primea şi a rămas la predicţia lui iniţială, spunând: „Nu, *ştiu* că sunt în Johannesburg – dacă mănânci pizza nu înseamnă că eşti în Italia”. Cu alte cuvinte, creierul său a modulat descendent semnalul de eroare. În consecinţă, domnul S. nu şi-a ajustat

programul de acţiune, fapt care nu avea cum să nu se termine într-un mod neplăcut – pentru că nu a reuşit să înveţe din experienţă. Dacă familia şi medicii nu s-ar fi îngrijit de toate nevoile sale, domnul S. cu siguranţă ar fi murit[39].

Majoritatea oamenilor nu îşi dau seama că percepţiile noastre de tip acum-şi-aici sunt ghidate permanent de predicţii, generate în principal din memoria pe termen lung. Dar aşa stau lucrurile. Din acest motiv, mult mai puţini neuroni propagă semnale de la organele senzoriale externe la sistemele mnezice interne decât invers[40]. De exemplu, raportul dintre conexiunile care intră şi cele care ies în corpul geniculat lateral (ce transmite informaţie de la ochi la cortexul vizual şi invers; vezi figura 6) este de aproximativ 1:10. Cea mai importantă contribuţie o au semnalele predictive care le *întâmpină* pe cele senzoriale ce sosesc de la periferie. Astfel se economiseşte un volum enorm de prelucrare a informaţiilor şi, prin urmare, travaliu metabolic. Având în vedere că creierul consumă cam 20% din rezervele noastre totale de energie, această eficienţă e preţioasă. De ce să tratezi orice lucru ca şi cum nu l-ai mai întâlnit niciodată? În schimb, creierul propagă spre interior numai acea parte a informaţiilor primite care *nu* se potriveşte cu aşteptările sale. Din acest motiv, astăzi percepţia este descrisă uneori drept „fantezie” şi „halucinaţie controlată”; începe cu un scenariu *aşteptat* care este apoi ajustat astfel încât să se potrivească cu semnalul primit[41]. În acest sens, anatomiştii clasici au avut dreptate: procesarea corticală constă în principal

din activarea unor „imagini mnezice”, rearanjate corespunzător astfel încât să prezică următorul ciclu de percepţie şi acţiune.

Afirmaţiile mele sunt valabile în cazul tuturor vertebratelor, nu numai al oamenilor. Poate că sunt valabile chiar în cazul tuturor organismelor înzestrate cu creier sau cu sistem nervos. (Insectele, de exemplu, au structuri cerebrale care funcţionează într-un mod similar cu sistemul nostru reticulat activator[42].) Ştim deja că forma fundamentală a conştiinţei este afectul, care ne permite să ne „simţim” calea de ieşire din situaţii neanticipate. Dar cum anume se transformă afectul endogen într-o exterocepţie conştientă?

Panksepp sugerează că o punte evolutivă între aceste două aspecte ale experienţei ar fi putut fi construită de „afectele senzoriale” (de exemplu, durerea, dezgustul şi uimirea). Afectele senzoriale sunt în acelaşi timp senzaţii interioare *şi* percepţii exterioare; ele sunt în mod inerent percepţii cu valenţe condiţionate de senzaţii specifice. Aşadar, de exemplu, durerea se simte diferit faţă de dezgust şi răspundem la ele în mod diferit, retrăgându-ne sau vomitând, în funcţie de cea pe care o resimţim. Conform lui Panksepp, pe această punte evolutivă conştiinţa s-a extins asupra percepţiei în general, întrucât contextualizează afectul. Până la urmă, lumea exterioară dobândeşte valoare doar pentru noi deoarece este locul unde trebuie să ne satisfacem nevoile; ea capătă valoare *în raport cu* afectul.

Aşadar, răspunsul meu la întrebarea de ce resimţim percepţia, acţiunea şi cogniţia este că *le resimţim*

pentru că ele contextualizează afectul. E ca şi cum experienţa noastră perceptivă spune: „Simt aşa *despre asta”*[43]*.* Cum s-ar spune, percepţia este incertitudine aplicată, afirmaţie ce justifică opinia lui Merker conform căreia conştiinţa afectivă şi cea perceptivă utilizează o „monedă comună” – ele sunt *tipuri* diferite de sentimente, dar sunt sentimente[44]. Aici vorbesc numai despre percepţia *conştientă.* Iar de aici nu mai e decât un pas mic până la a deduce că cele cinci modalităţi de conştiinţă perceptivă ale omului (văzul, auzul, pipăitul, gustul şi mirosul)[45] au evoluat pentru a descrie diferitele categorii de informaţii exterioare înregistrate de organele noastre senzoriale[46], aşa cum cele şapte tipuri de conştiinţă afectivă discutate în capitolul anterior descriu diferitele categorii de nevoi emoţionale la mamifere.

Dar este important să ne amintim şi că percepţia corticală este *apercepţie.* În conştiinţă nu apar semnalele senzoriale brute transmise de la periferie, ci *inferenţe* predictive derivate din urmele mnezice ale unor asemenea semnale şi ale consecinţelor lor. Faptul că abilitatea de a simţi se întinde, ca să zic aşa, peste inferenţele corticale stabilizate creând ceea ce eu numesc „corpuri solide psihice”[47] – lumea exterioară aşa cum se manifestă ea în percepţie – explică parţial calităţile fenomenologice diferite ale afectului versus percepţia conştientă. Voi reveni asupra acestor teme în capitolul 10.

Modelul interior al creierului este harta pe care o folosim pentru a naviga prin lume – chiar pentru *a genera* o lume preconizată. Dar nu ne putem lua drept

bune toate predicţiile. De fapt, modelul interior generează două aspecte ale „contextului preconizat”: pe de o parte, avem conţinutul propriu-zis al predicţiilor noastre, iar pe de altă parte, *nivelul de încredere* în acurateţea lor. Întrucât toate predicţiile sunt probabilistice, trebuie codificat şi gradul de *incertitudine preconizată* aferentă. Predicţiile sunt furnizate de reţelele memoriei pe termen lung din prozencefal, care filtrează prezentul prin prisma trecutului. Dar a doua dimensiune – ajustarea nivelurilor de încredere – este esenţa activităţii desfăşurate prin activarea modulatoare.

Asta s-a dereglat în cazul domnului S.: a pus prea multă bază în predicţiile lui. Cu alte cuvinte, din cauza afecţiunii sale, nu a acordat suficientă importanţă semnalelor de eroare (le-a acordat prea puţină *atenţie)*[48]*.* Cum e posibil acest lucru? Într- un fel, am analizat deja acest lucru din perspectiva cealaltă. Modularea semnalelor neuronale înseamnă pur şi simplu ajustarea în sus sau în jos a intensităţii lor. E ceea ce face triunghiul de decizie din mezencefal atunci când selectează ce semnale să intensifice – *modulează* intensitatea semnalului prin intermediul sistemului reticulat activator. Intensitatea iniţială a semnalului reprezintă contextul *preconizat.* Totuşi, pe măsură ce se dezvăluie contextul real (resimţit), intensităţile semnalelor trebuie ajustate. De aici decurge că încrederea sporită într-un semnal de eroare implică cu necesitate o încredere scăzută în predicţia ce a dus la eroare. Dacă ţi se declanşează alarma de fum, dar nu pare să ardă nimic, de vină sunt fie alarma, fie

aparenţele şi, cumva, trebuie dat un verdict în ce priveşte această neconcordanţă. Asta se întâmplă şi în triunghiul de decizie din mezencefal: se evaluează cererile concurente şi se declară un câştigător. Rezultatul acestui proces este transmis spre prozencefal de sistemul reticulat activator, care acţionează în funcţie de preconizările tale şi apoi – pe măsură ce acţiunea selectată se desfăşoară – eliberează nori de molecule neuromodulatoare ca să rafineze semnalele din reţelele memoriei tale pe termen lung, reglând pozitiv canalele prozencefalului în care sunt depozitate unele predicţii şi negativ altele. La rândul său, acest lucru duce la învăţarea din experienţă. Astfel, continuăm să ne îmbunătăţim modelul generativ al lumii, prin încercare şi eroare.

Tot ceea ce facem pe tărâmul incertitudinii este ghidat de aceste niveluri fluctuante ale încrederii. Credem că ştim ce se va întâmpla dacă acţionăm într-un anumit fel, dar oare chiar ştim? Dacă convingerea ne scade sub un anumit prag, atunci nu acţionăm sau schimbăm abordarea. În sfera exteroceptivă, treaba merge bine când se desfăşoară conform aşteptărilor şi prost când predomină incertitudinea. În consecinţă, *simţim* că treaba merge bine sau prost: creşterea încrederii (într-o predicţie) este bună, iar scăderea încrederii este rea. Deci încercăm să minimizăm incertitudinea în aşteptările noastre.

După cum am văzut mai sus, percepţia şi acţiunea sunt un proces continuu de *testare a ipotezelor,* în care creierul încearcă permanent să elimine semnalele de eroare şi să-şi confirme ipotezele. (Ştiinţa

experimentală în sine este doar o versiune sistematizată a acestui proces de zi cu zi, după cum am văzut în capitolul 1: „Dacă ipoteza X este corectă, atunci, când fac Z, ar trebui să se întâmple Y”.) Cu cât ţi se confirmă ipotezele, cu atât eşti mai încrezător şi cu atât mai puţin activat – mai puţin conştient – ai nevoie să fii. Îţi poţi automatiza secvenţele de acţiune, trecând în starea implicită. Dar, dacă ajungi într-o situaţie neprevăzută – pe care modelul tău predictiv nu pare să o clarifice într-un mod demn de încredere –, consecinţele acţiunilor tale devin predominante. Ieşi de pe pilot automat şi devii hiperconştient: triunghiul de decizie îţi ajustează cu grijă predicţiile pe măsură ce cauţi, în funcţie de ceea ce simţi, o cale de a evita consecinţele acţiunilor tale şi faci noi alegeri.

Astfel, simţirea, atât afectivă, cât şi cognitivă, rămâne factorul comun în orice formă de conştiinţă. Funcţia simţirii este să evalueze succesul sau eşecul programelor tale de acţiune şi al contextelor aferente acestora. Dar nu toate eşecurile sunt la fel de semnificative. Sentimentele cele mai puternice au legătură cu incertitudinile importante pentru supravieţuire, şi nu cu cele ataşate de posibilităţi relativ minore. Predicţiile proaste precum „Maşina pe care o urmăresc o va lua acum la stânga” pot fi gestionate la niveluri periferice din ierarhia predictivă, unde este tolerat un grad mai mare de eroare, şi, prin aceasta, determină schimbări în ceea ce numim atenţie, nu afect (deşi atenţia este modulată în egală măsură de trunchiul cerebral reticulat). Totuşi, dacă deodată creşte probabilitatea ca maşina pe care o urmăreşti să te implice într- un

accident, probabil că eroarea ta va activa un răspuns afectiv. Atenţia poate fi şi direcţionată, şi captată. În al doilea caz e însoţită de sentimente de uluire sau spaimă, ce rescriu din mers predicţiile eşuate, astfel încât modelul lumii să nu îţi mai facă aceeaşi surpriză neplăcută.

Iată cum funcţionează conştiinţa în creier, la nivel fiziologic.

Poate că nu pare cu totul satisfăcător. De pildă, nu prea mi-am bătut capul cu abordările filosofice ale problemei minte/corp. Nu m-am ocupat de relaţia dintre ceea ce unii filosofi numesc „aspectele ei duale”: de ce fiziologia obiectivă a conştiinţei este însoţită de o senzaţie fenomenologică subiectivă? Atunci când oamenii de ştiinţă observă o asemenea corelaţie regulată, între oricare două seturi de date, ei caută o singură cauză fundamentală. Vor *să explice de ce* fenomenele se întâmplă în acelaşi timp. Aşadar, ca să înţelegem legăturile regulate dintre aspectele fiziologice şi cele psihologice ale conştiinţei, trebuie să pătrundem mai în profunzime. Trebuie să depăşim constrângerile disciplinare ale fiziologiei şi psihologiei. De obicei, noi, oamenii de ştiinţă, facem acest lucru apelând nu la metafizică, ci la fizică[49]. În acest caz, răspunsurile de care avem nevoie trebuie căutate în fizica entropiei†[3].

† 1 De remarcat că unele molecule neuromodulatoare pot acţiona *şi* ca neurotransmiţători.

*2 În toată cartea, folosesc cuvintele „deasupra”, „dedesubt”, „mai sus”, „mai jos”, „ascendent”, „descendent” etc. nu ca judecăţi de valoare, ci ca

Dimpotrivă, din punct de vedere funcţional, structurile cele mai de sus ale prozencefalului sunt pur şi simplu *dezvoltări* ale structurilor celor mai de jos din trunchiul cerebral.

*3 Cititorii care nu au urmărit notele de final ar putea consulta acum anexa de la p. 313. Deşi e oarecum tehnică, ea face o trecere utilă spre capitolul următor.

localizatori anatomici. Spre deosebire de alte organe, creierul este ierarhic *din punct de vedere structural.* E stratificat oarecum asemenea unui sit arheologic, nivelurile mai vechi fiind acoperite de cele mai noi. De unde şi titlul acestei cărţi: Izvorul *ascuns.* Centrul cel mai profund al trunchiului cerebral conţine structurile cele mai vechi, din punctul de vedere al evoluţiei, iar nivelurile superioare ale cortexului le conţin pe cele mai recente. Asta nu înseamnă că structurile inferioare (şi mai vechi) sunt mai puţin importante decât cele superioare (mai noi).

Principiul energiei libere

Un grup de oameni de ştiinţă sunt într-un laborator, urmărind un monitor mare de calculator. Puncte şi pete se rotesc pe ecran. Se pot distinge diferite culori: albastru, roşu, violet şi altele. Şi punctele par să aibă mărimi diferite, dar mişcările lor haotice nu au un tipar. Se vălătucesc şi se întrepătrund ca nişte nori de gaz, umplând la întâmplare spaţiul virtual. Cu toate acestea, un ceas digital marchează trecerea timpului în secunde, iar două axe în partea de jos a ecranului, în stânga, identifică localizarea spaţială a fiecărui punct. Asta sugerează că se întâmplă ceva măsurabil. Dar cum se pot cuantifica mişcările în acest haos?

Îl întrebi pe unul dintre fizicienii din încăpere la ce ne uităm, iar el spune, fără ca răspunsul să te ajute, că este un proces „stocastic”. Stocastic înseamnă aleatoriu. Treptat, observi că punctele mai mici par să fie mai lente decât cele mai mari; par să danseze pe o melodie uşor diferită. La indicaţiile unui neurolog din încăpere, informaticianul scrie ceva la tastatură, iar vârtejurile de pe ecran încremenesc brusc. Informaticianul salvează atent datele. Apoi, tot la

cererea neurologului, tastează o mulţime de cifre care schimbă valorile într-un set de ecuaţii care acum apar pe un ecran mic de lângă cel mare la care priveşti. Îţi spune că ajustează „interacţiunile locale ale subsistemelor”.

Din nou, particulele ca nişte nori apar în colţul din stânga jos al ecranului şi încep să se rotească. De data asta, după o scurtă perioadă de haos, e mai uşor să discerni un tipar. Punctele colorate fac valuri în principal spre exterior la început, apoi, treptat, se adună în mase şi converg spontan spre centrul ecranului, formând o masă mai degrabă circulară. Cu puţină imaginaţie, ar putea fi un stol compact de păsări (poate grauri) care zboară coordonat. Constant, mişcarea particulelor devine mai ordonată. Se înghiontesc ca nişte soldaţi ocupându- şi poziţiile stabilite pentru o paradă, până ce începe să se întrevadă o structură clară: patru straturi concentrice. În centru sunt punctele albastru-închis. Acestea sunt înconjurate de cele roşii, care, la rândul lor, sunt înconjurate de cele violet. Punctele albastru-deschis formează un fel de limită exterioară. Toate acestea se petrec pe un fond de puncte negre mai mici care par să continue să se mişte haotic. Neurologul pare mulţumit. Îi cere informaticianului să oprească mişcările de pe ecran. Ceasul arată că au trecut 1.278 de secunde.

Ne aflăm la Centrul Wellcome pentru Neuroimagistică Umană din Queen Square, Londra, unde Karl Friston este directorul ştiinţific. Te-a invitat să urmăreşti un experiment interesant: o simulare a interacţiunilor pe distanţe scurte ce apar între diferite

subsisteme fizice atunci când sunt supuse unor forţe diverse.

Regulile ce guvernează aceste particule virtuale sunt la fel de variate ca şi cele care guvernează comportamentul moleculelor şi atomilor reali: tendinţe neintenţionate (dar nu la întâmplare) de atracţie şi de respingere. Evident, asemenea interacţiuni generează *ordine din haos.* Se consideră că astfel de ordonări spontane au avut loc atunci când a apărut viaţă din supa primordială. Totuşi, experimentul se desfăşoară într-un centru de neuroştiinţe cognitive, vizavi de Spitalul Naţional de Neurologie şi Neurochirurgie, şi nu se supără nimeni dacă te întrebi ce legătură au cu creierul aceste particule virtuale.

Conform explicaţiilor lui Friston, sistemele biologice precum celulele trebuie să fi apărut prin intermediul unor variante complexe ale aceluiaşi proces care a format sisteme mai simple „cu autoorganizare”, cum sunt cristalele dintr-un lichid, pentru că au în comun un anumit mecanism. Acest mecanism, spune Friston, este „minimizarea energiei libere” (pe care îl voi explica puţin mai jos). Toate sistemele cu autoorganizare, inclusiv tu şi eu, au în comun o sarcină fundamentală: să-şi continue existenţa. Friston crede că realizăm această sarcină prin minimizarea energiei noastre libere. Cristalele, celulele şi creierii – spune el – nu sunt decât manifestări tot mai complexe ale acestui mecanism fundamental ce se autoconservă[1]. Într-adevăr, atât de multe aspecte a ceea ce considerăm a fi viaţă psihică apar în zorii organizării biologice, încât contribuţia creierului propriu-zis poate să pară destul

de subtilă. Dar, atât timp cât nu pierdem din vedere conceptul de energie liberă, totul – chiar totul – va deveni clar.

Karl Friston este ultimul dintre marii oameni de ştiinţă care m-au influenţat din punct de vedere profesional. Îl consider un geniu şi (în mod obiectiv) specialistul în neuroştiinţe cel mai influent din lume la ora actuală. Influenţa se măsoară prin „indicele Hirsch”, ce măsoară impactul lucrărilor pe care le-ai publicat[2]. Ca regulă generală, dacă indicele tău Hirsch este mai mare decât numărul anilor de când ţi-ai obţinut doctoratul, atunci stai bine. Indicele Hirsch al lui Friston este 235, mai mare decât al oricărui alt specialist în neuroştiinţe[3]. A devenit celebru cu „reprezentarea statistică parametrică”, ce a permis analiza neuroimagisticii funcţionale atât de răspândită astăzi. Totuşi, activitatea sa în ce priveşte „codificarea predictivă” şi principiul energiei libere i-au adus mai multă recunoaştere.

În pofida reputaţiei extraordinare a lui Friston, timp de mulţi ani nu prea m-am interesat de activitatea sa. Apoi, în 2010, a publicat o lucrare împreună cu un tânăr psihofarmacolog pe nume Robin Carhart-Harris, de care ştiam că e interesat de neuropsihanaliză. Lucrarea acestor doi autori argumenta că noţiunea de energie pulsională (adică „energie psihică”) a lui Freud corespunde principiul energiei libere[4]. După cum am explicat mai sus, Freud a admis imediat că „aici s-a petrecut ceva complet insesizabil pentru noi”, cu referire la modul în care nevoile corpului ar putea

deveni energie psihică. A scris şi că această energie se pretează la creştere, reducere, dislocare şi descărcare, având astfel toate caracteristicile unei cantităţi, „deşi nu avem mcmn mijloc prin care să o măsurăm”[5]. Ţinând cont că intenţia iniţială (abandonată) a lui Freud fusese de a „prezenta procesele psihice ca pe nişte stări determinate cantitativ”, lucrarea lui Carhart-Harris şi Friston a fost ca o lovitură de trăsnet pentru mine. Dacă energia mintală chiar este izomorfă cu schimbările energiei libere termodinamice, atunci – m-am gândit eu – trebuie să fie şi măsurabilă, şi reductibilă la legile fizicii.

Prin urmare, am studiat lucrările anterioare ale lui Friston şi l-am căutat. Ne-am întâlnit de mai multe ori în anii ce au urmat, la Londra şi la Frankfurt. Principalul subiect al conversaţiilor noastre a fost rolul afectului în viaţa psihică. Deoarece activitatea lui Friston din perioada aceea era legată preponderent, ca în cazul mai tuturor specialiştilor, de cortex, mecanismele predictive pe care le descoperise vizau aproape exclusiv *cogniţia.* Din acest motiv, de exemplu, un celebru articol al său ce arată că codificarea predictivă explică modul în care neuronii comunică între ei are titlul „A theory of *cortical* responses”[6]. Trebuie să mărturisesc însă că nu înţelesesem pe deplin multe dintre articolele sale mai tehnice.

În 2017, Friston a fost invitat ca vorbitor principal la congresul nostru anual de neuropsihanaliză (ce s-a ţinut în acel an în vechiul University College Hospital, Londra, având ca temă principală codificarea

predictivă). Ca să-mi formulez observaţiile finale fără să mă fac de râs, trebuia să stăpânesc chestiunile de fizică. Prin urmare, printre multe alte lucrări ale lui Friston, i-am recitit un articol foarte tehnic apărut în una dintre revistele publicate de Royal Society. Se numea „Life aş we know it”[7]. Cu mare efort, l-am înţeles corect pentru prima dată. Acest articol nu îşi propunea nimic altceva decât să reducă la nişte ecuaţii matematice legile fundamentale ce guvernează *intenţionalitatea.*

Implicaţiile erau electrizante. Mi s-a părut că aceste ecuaţii îmi puteau oferi pasul înainte pe care îl căutam. Aşa că, imediat după discuţiile noastre ştiinţifice de la congresul din 2017, i-am scris propunându-i să ne punem ideile laolaltă, ca să încercăm să încorporăm conştiinţa în principiul energiei libere. Spre încântarea mea, Friston mi-a acceptat propunerea şi am început să colaborăm la un articol în care am schiţat ceea ce a devenit opinia noastră comună[8].

Legătura dintre activitatea lui Friston şi cercetările mele este *homeostaza.* Am explicat mai sus că trebuie să ne menţinem între limitele noastre viabile din punct de vedere fiziologic. Să folosim termoreglarea ca exemplu. Nu te poţi mulţumi cu orice temperatură corporală: trebuie să te menţii în intervalul limitat 36,5-37,5°C. Dacă te încălzeşti semnificativ mai mult, mori; dacă te răceşti semnificativ mai mult, mori. Nu poţi permite temperaturii corporale să devină egală cu cea a mediului ambiant, cum face apa fierbinte când o adaugi într-o cadă cu apă rece. Apa fierbinte care curge în cadă nu rămâne separată de cea rece, într-o globulă mare, sub robinet. Însă *tu* da – şi aşa şi trebuie, ca să rămâi în

viaţă –, iar asta presupune *travaliu.* Pacienţii comatoşi nu pot realiza un asemenea travaliu, aşa că mor din cauze precum hipertermia; literalmente, se supraîncălzesc[9]. Acelaşi lucru e valabil şi în cazul reglării gazelor sangvine, al echilibrului fluidelor şi al celui energetic şi al multor altor procese fiziologice. E valabil chiar şi în cazul nevoilor *emoţionale,* care, după cum am văzut în capitolul 5, nu sunt mai puţin „biologice” decât cele corporale. Şi menţinerea în limitele viabile ale emoţiilor noastre presupune travaliu: ca să rămânem aproape de persoanele care au grijă de noi, ca să scăpăm de prădători, ca să îndepărtăm obstacole şi aşa mai departe. Dincolo de un anumit nivel de predictibilitate, travaliul necesar ca să facem aceste lucruri este reglat de sentimente.

Mecanismul pe care tocmai l-am descris este o formă extinsă de homeostază şi nu e complicat (vezi figura 12).

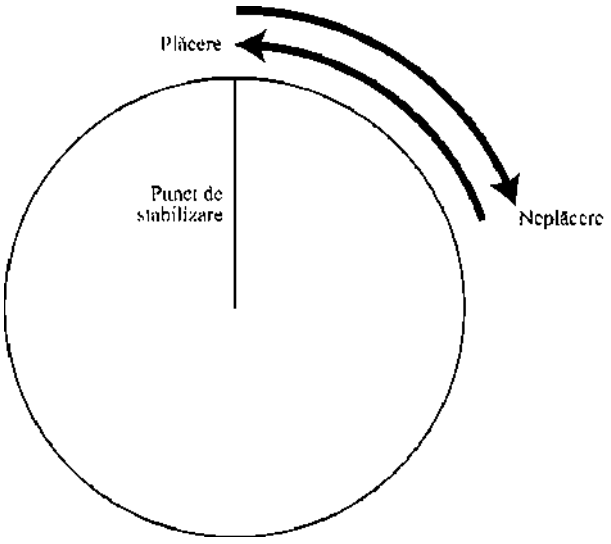


Figura 12. Homeostaza sentimentului. Punctul de stabilizare

reprezintă limitele viabile ale sistemului.

Fiecare homeostat constă din doar trei componente: un *receptor* (ce măsoară temperatura, în exemplul meu), un *centru de control* (ce determină modul în care se menţine temperatura în limitele viabile: 36,5-37,5°C în exemplul meu) şi un *efector* (care face munca necesară pentru a te readuce în limitele respective când le depăşeşti). Pentru că mecanismul homeostazei este atât de simplu, poate fi redus la legi ale fizicii. Despre asta este vorba în articolul lui Friston: legile fundamentale ce guvernează „viaţa aşa cum o ştim”.

M-am entuziasmat când am realizat că aceste legi – extinse astfel încât să includă forma mai predictibilă de homeostază ce stă la baza afectului – ar putea să explice şi conştiinţa: şi nu numai elementele observabile exterioare ale comportamentului conştient *(fiziologia* obiectivă a triunghiului de decizie din mezencefal şi a sistemului reticulat activator), ci şi cele observate în interior. Aceasta promitea să explice *sentimentele* subiective care ghidează deciziile.

În acest punct, drumul meu şi cel al lui Damasio s-au despărţit. Poate mai mult decât ceilalţi specialişti în neuroştiinţe, el atrăsese atenţia asupra faptului că afectul (şi, prin urmare, conştiinţa) este la bază o formă de homeostază. Dar a respins inferenţa mea suplimentară: în cazul în care conştiinţa este homeostatică, iar homeostaza este reductibilă la legile fizicii, atunci şi fenomenele conştiinţei pot fi reduse la legi din fizică[10]. Mecanismul conştiinţei, nu mai puţin

decât cel al mişcării corpurilor cereşti şi a orice altceva din natură, *trebuie* să poată fi explicat în baza unor legi şi, prin urmare, să fie oarecum predictibil, chiar dacă doar la un nivel probabilistic. Când Damasio a citit o variantă preliminară a articolului pe care l-am scris împreună cu Friston, am avut o discuţie tensionată în biroul lui. Nu putea să înţeleagă de ce încercam să reduc conştiinţa la ceea ce el numea „algoritmi”. Era cazul clasic al specialistului ce refuză implicaţiile unei idei pe care el însuşi a lansat-o în lume – ca răspunsul lui Einstein la mecanica cuantică: „Dumnezeu nu joacă zaruri cu universul”[11].

I-am expus pe scurt lui Damasio argumentaţia pe care o voi explica în cele ce urmează[12]. Totuşi, ca să explic cum trebuie, este necesar să prezint fizica pe care am fost nevoit să o stăpânesc ca să înţeleg principiul energiei libere. Trage adânc aer în piept.

Esenţa homeostazei este că organismele vii trebuie să ocupe un număr limitat de stări fizice: stările lor viabile ori stările lor valoroase sau preferate ori ceea ce Friston numeşte (referindu- se la toate acestea) stările lor „preconizate”. Nu ne putem permite să ne risipim în toate stările posibile. S-a dovedit că acest imperativ biologic are o legătură profundă cu unul dintre conceptele explicative fundamentale din fizică, şi anume *entropia.* Majoritatea oamenilor înţeleg într-un mod intuitiv ce este „entropia”. O concep ca pe o tendinţă naturală spre dezordine, disipare, dezintegrare şi altele asemenea. Legile entropiei sunt ceea ce face gheaţa să se topească, bateriile să se descarce, bilele de biliard să

se oprească, iar apa fierbinte să se amestece cu cea rece.

Homeostaza merge în direcţia opusă. *Se opune* entropiei. Se asigură că ocupi un număr limitat de stări posibile. În acest fel, îţi menţine temperatura necesară şi te ţine viu – şi nu te lasă să te dezintegrezi. Fiinţele vii *nu au cum să nu* se opună unuia dintre principiile fundamentale ale fizicii: a doua lege a termodinamicii.

Prima lege a termodinamicii se referă la conservarea energiei[13]. Conform acesteia, energia nu poate fi creată sau distrusă, ci poate fi doar transformată dintr-un tip într-altul şi poate trece dintr-un loc într-altul. (Mai ştim, datorită lui Einstein, şi că poate fi transformată în materie.)

Conform celei de-a doua legi, procesele naturale sunt întotdeauna ireversibile[14]. Astfel, apa fierbinte din cadă amestecată cu apa rece nu poate fi separată ulterior. De asemenea, energia din căldură nu poate fi pusă la loc în cărbunele ars care a produs-o şi nici energia pierdută pe parcurs. Acest lucru se datorează entropiei, care, în consecinţă, *creşte mereu* la scara universului[15]. De fapt, s-ar putea ca entropia să fie baza fizică a faptului că însuşi timpul pare că are o direcţie şi o curgere.

În termodinamică, există două stări ale energiei: exergică şi nonexergică. „Exergicitatea” energiei este definită prin capacitatea ei de a efectua lucru mecanic. De exemplu, energia dintr-o bucată de cărbune poate fi arsă pentru a produce căldură[16], care poate fierbe apă pentru a produce abur, care poate pune în mişcare un motor. Dar în fiecare etapă a acestui proces se pierde o

anumită cantitate de energie. Adică nu o poţi folosi niciodată în întregime într-un mod util pentru lucru mecanic.

Coroborând aceste fapte, aflăm următoarele: pe măsură ce exergia dintr-un sistem se epuizează, entropia creşte. În consecinţă, capacitatea sistemului de a efectua lucru mecanic scade întotdeauna. Prin urmare, entropia este asociată cu pierderea de exergie, din cauză că energia nu mai este disponibilă pentru a efectua lucru mecanic. Legea a doua este o afirmare a faptului ineluctabil că o parte din energie se pierde în lucru mecanic util în timpul oricărui proces natural[17].

Am menţionat o anumită înţelegere intuitivă a „entropiei” puţin mai sus, dar definiţia formală, tehnică a entropiei se referă la numărul de stări distincte pe care le poate ocupa un sistem[18]. Entropia este determinată de numărul de stări microscopice posibile care ar duce la aceeaşi stare macroscopică. Mai simplu spus: *cu cât stările posibile sunt mai puţine, cu atât e mai scăzută entropia.*

Homeostaza stabileşte limite ale tipurilor de stări macroscopice pe care le poate ocupa un sistem ca mine sau ca tine. Aminteşte-ţi că homeostaza ne menţine în viaţă efectuând un lucru mecanic; prin urmare, dacă entropia presupune pierderea capacităţii de a efectua lucru mecanic, ea e „rea” pentru noi, ca sisteme biologice. Funcţia elementară a fiinţelor vii este să se opună entropiei.

Exemplul utilizat în mod convenţional de fizicieni

pentru a ilustra aceste concepte implică un gaz comprimat introdus într- o cameră mai mare, goală, printr-un orificiu mic. Pe măsură ce moleculele se mişcă aleatoriu, ele explorează camera, răspândindu-se pentru a ocupa spaţiul disponibil. Cu cât trece timpul, cu atât fiecare moleculă se poate găsi în mai multe locuri. Singura modalitate prin care poate fi inversat acest proces – singura modalitate prin care gazul poate fi readus în recipientul său iniţial – presupune lucru mecanic.

Gândeşte-te la entropie prin raportare la numărul de locuri în care s-ar putea afla fiecare moleculă la un moment dat. Aceasta se dovedeşte a fi o afirmaţie despre *probabilitate.* Probabilitatea statistică a fiecărei molecule de a ocupa o anumită poziţie se reduce pe măsură ce creşte entropia: pe măsură ce gazul se dilată, poziţia fiecărei molecule devine mai puţin predictibilă. Creşterea entropiei înseamnă *reducerea predictibilităţii.*

Acest lucru este important deoarece, spre deosebire de celelalte legi ale termodinamicii, legile probabilităţii se aplică tuturor lucrurilor, nu numai celor materiale. Aşa cum entropia unui gaz dintr-o cameră poate fi definită sub aspect probabilistic, tot astfel şi entropia poate fi asociată cu un proces psihologic de luare a deciziilor. În ambele cazuri, entropia creşte odată cu caracterul aleatoriu al rezultatelor posibile. „Entropia” asociată cu gazele ce se dilată şi cea asociată cu opţiunile al căror număr creşte *sunt unul şi acelaşi lucru.* Nu toate lucrurile care există în natură sunt vizibile şi tangibile, dar toate se supun legilor

probabilităţii. Acesta este motivul pentru care probabilitatea tulbură fizica modernă în esenţa ei, unde materia nu mai este considerată un concept fundamental, iar particulele clasice dispar[19].

În exemplul fizic al entropiei, când un gaz a fost iniţial comprimat în containerul său, iar moleculele sale erau înghesuite, pentru a descrie locul fiecărei molecule au fost necesari mai puţini *biţi de informaţie* decât după ce gazul a fost eliberat pentru a umple spaţiul disponibil în camera mare. În informatică, o „cifră binară” – pe scurt, „bit” – este unitatea elementară de informaţie. Un bit poate să aibă una din două valori opuse – de exemplu, da versus nu, pornit versus oprit, pozitiv versus negativ. De obicei, aceste stări sunt reprezentate ca 1 versus 0[20].

Iniţial, sunt necesari mai puţini biţi de informaţie pentru a descrie gazul, dar, pe măsură ce acesta se dilată, numărul de stări posibile pe care le poate ocupa fiecare moleculă creşte. Astfel, entropia (măsurată de obicei din punct de vedere fizic în termeni de căldură schimbată şi de temperatură) poate fi măsurată şi în biţi: cu cât e necesară mai multă informaţie pentru a descrie microstarea unui sistem (adică starea fiecărei molecule), cu atât creşte entropia termodinamică. Sau, în termeni şi mai simpli: *cu cât e mai mare numărul de întrebări cu răspunsuri de tip da/nu la care trebuie să se răspundă pentru a descrie un sistem, cu atât mai mare este entropia lui.* Astfel, când microstările unui sistem au valori de probabilitate scăzute, fiecare măsurare a sistemului conţine mai multă informaţie decât atunci când ar avea o

probabilitate mare – fiindcă ar trebui să se răspundă la mai multe întrebări binare pentru a descrie totalitatea stării sistemului.

Entropia este minimă când răspunsul la fiecare întrebare implicând răspunsuri de tip da/nu este cu totul predictibil, adică atunci când nu se află nimic şi nu se obţine nicio
informaţie. Conţinutul informaţional al datului cu banul este un
bit, deoarece şansele să iasă cap sunt de 50%, dar obţii zero informaţie când foloseşti un ban cu două capete, deoarece şansele să iasă capul sunt de 100%. Nu furnizează nicio informaţie pentru că răspunsul este în întregime predictibil. Entropia măsoară cantitatea medie de informaţie pe care o obţii prin măsurarea unui sistem de mai multe ori. Astfel, entropia unei serii de măsurători este *informaţia sa medie, incertitudinea sa* medie.

Ca să ne dăm seama de importanţa acestui lucru pentru o perspectivă neuroştiinţifică asupra conştiinţei, să ne amintim că tiparele cu unde lente sincronizate dintr-o electroencefalogramă (EEG) sunt mai ordonate (mai predictibile) decât cele rapide desincronizate (haotice). Prin urmare, tiparele „cu activare scăzută” conţin mai puţină informaţie decât cele „cu activare mare” (vezi figura 10). *Cele cu activare mai mare conţin mai multă incertitudine*[21]*.* Astfel, valorile entropiei din EEG sunt mai mari la pacienţii aflaţi într- o stare minimă de conştienţă decât la cei aflaţi în stare vegetativă[22]. Iar asta este explicabil: activitatea corticală din creierul conştient comunică mai multă

informaţie decât în timpul somnului profund. Iar acum vine partea ciudată: dacă mai multă informaţie înseamnă mai multă incertitudine şi, prin urmare, mai multă entropie, atunci – întrucât fiinţele vii trebuie să se opună entropiei – activitatea în stare de veghe este mai puţin dezirabilă, din punct de vedere biologic, decât somnul profund[23]. Ştiu că această afirmaţie este contrară logicii, dar va deveni mai uşor de înţeles pe măsură ce avansăm cu explicaţiile—.

Relaţia dintre entropie şi informaţie a fost formalizată într-o ecuaţie celebră de Claude Shannon, inginer electronist şi matematician. Cu această descoperire, Shannon a încorporat „informaţia” în fizică, în cadrul căreia a devenit de atunci un concept fundamental, mai ales în mecanica cuantică[25]. Pornind de la ideile lui Shannon, fizicianul Edwin Thompson Jaynes a argumentat că entropia termodinamică trebuie considerată o *aplicaţie* a entropiei informaţionale[26]. Prin urmare, definiţia lui Shannon este şi mai importantă decât cea din termodinamică: o definiţie abstractă a entropiei în termeni de dinamică a informaţiei este mai aplicabilă la un nivel general decât una concretă, în termeni de dinamică a căldurii. Aşadar, legile termodinamicii pot fi considerate *un caz special al legilor mai profunde ale probabilităţii.* Acest lucru este important deoarece legile termodinamicii se aplică numai sistemelor materiale (tangibile, vizibile), precum creierul, iar legile informaţiei se aplică şi sistemelor imateriale (intangibile, invizibile), precum psihicul.

Dar probabilitatea nu este chiar totuna cu

informaţia. Din perspectiva lui Shannon, informaţia presupune un factor suplimentar, *comunicarea*. De unde şi titlul articolului său fundamental care a consacrat cibernetica: „Teoria matematică a comunicării”[27]. Spre deosebire de probabilităţi – care există în sine –, comunicarea presupune atât o *sursă* a informaţiei, cât şi un *receptor* al informaţiei. (Nu e neapărat necesar ca comunicatorul să fie o persoană. Ar putea fi o carte, de pildă, sau orice sistem care deţine informaţie din care receptorul poate să înveţe.)

Acest lucru pune probleme mari oricărei ipoteze teoretice conform căreia conştiinţa *pur şi simplu este* informaţie[28], apărând întrebarea: care este sursa şi care este receptorul informaţiei – integrată sau de alt tip? Acesta este motivul pentru care eu nu sunt mulţumit de modelele fluxului de informaţii folosite de cognitivişti. Ei evită întrebarea „Unde este *subiectul,* receptorul?”. În acest fel, ca să-l parafrazez pe Oliver Sacks, psihicul este exclus din ştiinţele cognitive.

Această omitere a subiectului ce are experienţe duce la o întrebare şi mai importantă totuşi – poate cea mai importantă dintre cele cu care se confruntă astăzi ştiinţele cognitive: înainte de toate, în lipsa unui observator, cum şi de ce se întâmplă prelucrarea informaţiei (adică formularea unor întrebări, apoi a unor răspunsuri la ele)?

Descoperirea de către Shannon a informaţiei-ca-entropie l-a determinat pe fizicianul John Wheeler să propună o interpretare „participativă” a universului[29]. Conform lui Wheeler, lucrurile apar în forma în care apar (cu alte cuvinte, ca fenomene observabile) numai

ca răspuns la întrebările pe care le punem. Fenomenele ca atare există numai în ochii celui care le priveşte, ai unui observator participant, ai celui care pune întrebări. În celebra formulare a lui Wheeler, „cevaurile apar din biţi” (unde „cevaurile” sunt *lucruri* observabile)[30]. Aşadar, informaţia este „fizică” nu numai pentru că e implicată în legile fizicii, ci şi pentru că e baza tuturor *fenomenelor observabile.* Aşa devin observabile şi măsurabile forţele şi *r r r r* energiile abstracte: „După ce am luat tot în calcul, ceea ce numim realitate apare din adresarea întrebărilor cu răspunsuri de tip da/nu şi din consemnarea răspunsurilor evocate de echipament; pe scurt, [...] tot ce este fizic are origini informaţional-teoretice”—.

Modalităţile senzoriale ale sistemului nervos generează „răspunsuri evocate de echipament” la întrebările noastre despre univers. Răspunsurile senzoriale dau naştere fenomenelor – „lucrurilor” – pe care le trăim. Aşadar, experienţa însăşi apare din comunicarea dintre un receptor al informaţiei (un observator participant) şi o sursă a informaţiei, dintre ceva/cineva care pune întrebări şi răspunsurile pe care le consemnează. Dar tot nu am răspuns la o întrebare: *de unde vine acel ceva/cineva care pune întrebările?*

Chestiunile astea sunt cam complicate. Înainte să continuăm, propun să facem o recapitulare. Am expus trei idei importante. Prima este că informaţia medie a unui sistem este entropia sistemului respectiv (adică entropia dintr-un sistem indică cantitatea de informaţie necesară pentru a-i descrie starea fizică). A doua este că sistemele vii trebuie să se opună entropiei. Aceste

două fapte înseamnă că *trebuie să minimizăm informaţia pe care o procesăm.* (Aici mă refer la informaţie în sensul întâlnit la Shannon, desigur; cu alte cuvinte, trebuie să ne minimizăm *incertitudinea.)* Tot ce voi mai spune în acest capitol şi în următoarele două pleacă de la această concluzie simplă, dar uimitoare.

Ajungem astfel la a treia idee importantă pe care am aflat-o până acum: noi, sistemele vii, ne opunem entropiei prin mecanismul homeostazei. Pe scurt, primim informaţie despre supravieţuirea noastră probabilă *punând întrebări* (adică măsurând) despre starea noastră biologică în contextul unor evenimente aflate în desfăşurare. Cu cât răspunsurile sunt mai incerte (deci cu cât conţin mai multă informaţie), cu atât mai rău pentru noi – asta înseamnă că nu ne îndeplinim obligaţia homeostatică de a ocupa stări limitate (stările noastre preconizate).

Natura întrebărilor pe care le punem este determinată parţial de *specia* căreia îi aparţinem. Rechinii pot să respire în apă, oamenii nu. Aşadar, avem nevoi diferite şi ne aşteptăm să ocupăm stări diferite. Asemenea cerinţe sunt determinate prin selecţie naturală. Homeostaza înseamnă să rămâi în nişa în care te-a adus evoluţia speciilor. De aceea, fiecare specie trebuie să-şi pună întrebări precum „Pot să respir aici?”. Supravieţuirea noastră depinde de răspunsurile pe care le primim.

Apropo, de ce ar trebui să ne gândim la necesităţile biologice ca la nişte *aşteptări?* Această exprimare poate părea ciudată în acest punct, dar indică o continuitate profundă despre care vom vedea mai jos că e

importantă. Dacă te ajută, încearcă să priveşti din perspectiva evoluţiei înseşi, şi nu a unei creaturi individuale. Selecţia naturală a înzestrat fiecare specie pentru nişa sa ecologică: supravieţuirea fiecărei creaturi depinde numai de lucrurile care se găsesc de fapt fără probleme în habitatul său natural. Deci avem nevoie de aer *deoarece* ne putem aştepta să existe.

Acum pot să revin la întrebarea profundă formulată mai sus: de unde vin observatorii participanţi? Cu alte cuvinte: cum şi de ce, în termeni fizici, apare adresarea unor întrebări?[32]

Iată un foarte scurt istoric al ideii de autoorganizare. Primul care a utilizat acest termen este Immanuel Kant, în opera sa din 1790 *Critica puterii de judecare.* Kant afirmă că fiinţele au „scopuri” şi „obiective” intrinsece care pot fi adevărate numai dacă mecanismele lor componente sunt simultan atât scopuri, cât şi mijloace. Asemenea entităţi „teleologice” (adică entităţi cu scopuri şi obiective intrinsece), spune Kant, trebuie să aibă comportament intenţionat: „Numai în aceste condiţii şi în aceşti termeni poate un asemenea produs să fie o fiinţă organizată şi autoorganizată şi, ca atare, poate fi numită scop fizic”. Kant considera că modul în care asemenea fiinţe ar putea să apară e dincolo de puterea explicativă a ştiinţei: nu putea niciodată să existe „un Newton al firului de iarbă”.

Apoi Darwin a descoperit selecţia naturală. După cum ştim acum, selecţia naturală dă naştere obiectivelor şi scopurilor intrinsece ale *supravieţuirii* şi *reproducerii.* Acestea se dovedesc a fi amândouă

manifestări ale autoorganizării[33]. Odată cu ideea lui Darwin, chestiunea originii şi alcătuirii fiinţelor teleologice a devenit încadrabilă în domeniile ştiinţei[34]. Nu a mai rămas decât să li se desluşească detaliile.

Un alt pas important a fost făcut la mijlocul secolului XX, când Norbert Wiener, matematicianul care a fost întemeietorul „ciberneticii”, a adăugat ideea de *feedback* la perspectiva lui Shannon asupra informaţiei. Conform lui Wiener, un sistem îşi poate atinge obiectivul („starea de referinţă”) dacă primeşte feedback despre consecinţele acţiunilor sale. Feedbackul include semnalele de eroare – ce măsoară *deviaţiile* de la starea de referinţă – care ar fi folosite pentru ajustarea acţiunilor sistemului, astfel încât să-şi menţină cursul. Astfel, homeostaza se dovedeşte a fi un caz specific al unui principiu cibernetic mai general: e un fel de feedback negativ.

William Ross Ashby a folosit această idee de feedback, în combinaţie cu fizica statistică prezentată mai sus, ca să arate modul în care autoorganizarea se desfăşoară natural[35]. Ashby a arătat că multe sisteme dinamice complexe evoluează automat spre un *punct de stabilizare,* pe care l-a descris ca pe un „punct de atracţie” într-un „bazin” de stări învecinate. Evoluţia ulterioară a unor asemenea sisteme tinde apoi *să ocupe stări limitate.*

Sper că această tendinţă de a ocupa stări limitate îţi sună familiar: nu este nimic altceva decât o tendinţă de a te opune entropiei. Conform lui Friston, tocmai această tendinţă declanşează forme tot mai elaborate

de autoorganizare. După ce a creat această posibilitate între subsistemele din supa primordială pe care a simulat-o, descrisă la începutul acestui capitol, le-a observat comportamentul desfăşurându-se în trei etape:

1) în cazul unor parametri pe distanţe scurte, punctele doar au ţâşnit în toate direcţiile;

2) în cazul altor parametri, punctele s-au adunat în structuri stabile asemănătoare unor cristale;

3) iar în cazul altora, au manifestat comportamente mai complexe: după ce s-au adunat, s-au înghesuit neobosite unele în altele, ocupând poziţii specifice în cadrul unei structuri dinamice.

Iată descrierea făcută de Friston (nu-ţi bate capul cu limbajul tehnic, vreau doar să transmit o impresie vizuală a ceea ce a văzut):

> Aceste comportamente variază de la comportamente ca de gaz (unde subsistemele uneori se apropie suficient încât să interacţioneze) la un focar de activitate, când subsistemele sunt nevoite să se întâlnească în partea de jos a sursei potenţiale. În acest regim, subsistemele se apropie suficient de mult încât legea inversului pătratului să le distrugă – amintind de coliziunile particulelor subatomice în fizica nucleară. La anumite valori ale parametrilor, aceste evenimente sporadice şi critice pot să facă dinamica nonergodică, cu fluctuaţii impredictibile de mare amplitudine care nu se stabilizează. În alte regimuri apare o structură mai cristalină, cu interacţiuni mute şi entropie structurală

(configurativă) scăzută. Totuşi, în cazul celor mai multe valori ale parametrilor, comportamentul ergodic apare pe măsură ce ansamblul se apropie de punctul său de atracţie global aleatoriu (de obicei, după aproximativ 1.000 s): în general, iniţial subsistemele se resping reciproc (în mare parte ca în ilustrările Big Bangului), apoi cad înapoi spre centru, găsindu-se unul pe altul în timp ce se adună. Apoi, interacţiunile locale mediază o reorganizare, în care subsistemele sunt trecute dintr-o parte în alta (uneori spre periferie) până când vecinii se îngrămădesc uşor unul în altul. În ce priveşte dinamica, sincronizarea temporară poate fi văzută ca nişte valuri de izbucnire dinamică [...]. Pe scurt, mişcarea şi dinamica electrochimică seamănă foarte mult cu o supă în clocot (nu cu mult deosebită de erupţiile solare de pe suprafaţa soarelui) – dar are ea oare mai multă autoorganizare de-atât?[36]

Răspunsul la ultima lui întrebare s-a dovedit a fi pozitiv. Apare o structură dinamică complexă în care subsistemele dense, după ce s-au separat de mediul lor înconjurător, formează *structuri stratificate concentric,* fiecare având un centru interior şi o suprafaţă exterioară şi fiind împărţită suplimentar în două substraturi (vezi figura 13). Substraturile suprafeţei împărţite manifestă tipare foarte interesante de interacţiune cu centrul interior şi, respectiv, cu mediul înconjurător. Stările substratului exterior sunt influenţate de cele ale mediului exterior şi ele, la rândul lor, le influenţează pe cele ale subsistemelor interne, dar această influenţă nu este reciprocă (altfel spus,

componentele interne ale nucleului nu au niciun impact asupra substratului exterior). De asemenea, stările substratului interior sunt afectate cauzal de cele ale nucleului interior, iar ele, la rândul lor, le influenţează pe cele ale mediului exterior, dar, din nou, influenţa nu este reciprocă. Acest aranjament al dependenţelor cauzale defineşte proprietăţile a ceea ce se numeşte *zonă Markov*[37].

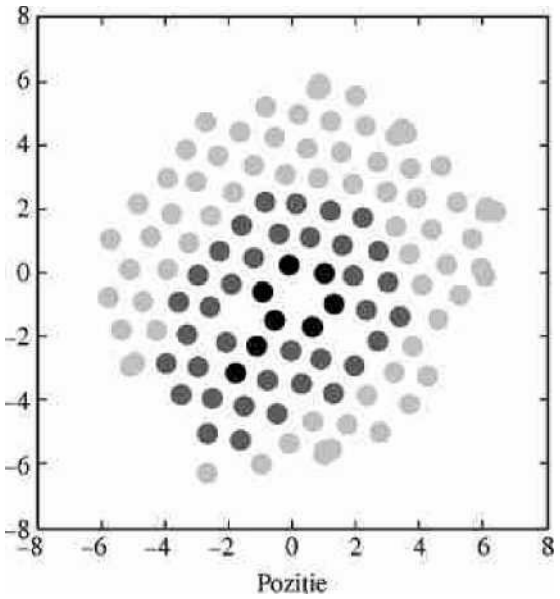


Figura 13. Un sistem cu autoorganizare, cu zona Markov
aferentă. În această imagine, nucleul interior al sistemului este
ilustrat prin punctele negre, iar straturile sale înconjurătoare
prin cele de culoare gri-închis: zona. Punctele gri-deschis sunt
exterioare sistemului. (Reprezentarea originală a lui Friston

diferenţia substraturile zonei. De precizat, întrucât culorile lui
sunt menţionate în text: punctele exterioare erau albastru-
deschis, cele din interior erau albastru-închis, cele senzoriale
erau violet, iar cele active erau roşii.)

„Zona Markov” este un concept statistic ce separă două seturi de stări. Asemenea formaţiuni induc o *împărţire* a stărilor în stări interne şi externe, adică într-un sistem şi un nu- sistem, astfel încât stările interne sunt izolate de cele externe sistemului. Cu alte cuvinte, stările externe pot fi „simţite” numai *indirect* de cele interne ca stări ale zonei. Mai mult, o zonă Markov este ea însăşi împărţită în subseturi care sunt şi în subseturi care nu sunt dependente cauzal (direct) de stările setului exterior. Aceste stări ale zonei sunt numite stări „senzoriale” şi, respectiv, „active”.

Astfel, formarea unei zone Markov împarte stările unui sistem în patru tipuri – interne, active, senzoriale şi externe –, unde stările externe nu fac parte din entitatea cu autoorganizare. Un aspect crucial, dependenţele dintre aceste patru tipuri de stări creează o cauzalitate circulară. Stările externe le influenţează pe cele interne prin stările senzoriale ale zonei, iar stările interne se conectează iar la cele externe prin stările lor active. În acest fel, stările interne şi cele externe se cauzează reciproc într-un mod circular. Altfel spus, stările senzoriale *furnizează feedback* despre consecinţele efectului asupra stărilor externe ale

stărilor active, ajustând astfel acţiunile ulterioare ale sistemului.

Dacă asta seamănă cu ciclul percepţie/acţiune la organismele vii – şi îţi aminteşte de ce am spus în capitolul anterior despre cauzalitatea circulară dintre sistemul reticulat activator, prozencefal şi triunghiul de decizie din mezencefal – nu e o întâmplare. Exact asta este valoarea unor astfel de modele abstracte. Relevă formalisme profund regulate care pot fi apoi recunoscute în substraturi de multe tipuri – şi ne permit să înţelegem din perspective noi structura organismelor vii.

Odată ce începi să cauţi, vezi zone Markov peste tot. O membrană celulară are proprietăţile unei zone Markov, la fel şi pielea şi sistemul musculoscheletal al corpului în ansamblu – care este el însuşi alcătuit din celule. La fel şi fiecare organit, organ şi sistem fiziologic. Aşadar, creierul (de fapt, întregul sistem nervos) – ce reglează celelalte sisteme ale corpului – are o zonă Markov. De fapt, este o *metazonă,* deoarece le înconjoară pe toate celelalte. Sistemele cu autoorganizare pot fi întotdeauna alcătuite din sisteme cu autoorganizare mai mici – nu până la ultimul element, dar cu siguranţă până la unele chinuitor de mici. Aceasta este materia primară a vieţii: miliarde de homeostate mici înfăşurate în zonele lor Markov.

Ne apropiem de un răspuns la întrebarea pe care am pus-o mai sus: de ce şi cum, în termeni fizici, apare formularea unor întrebări? Încă nu am ajuns acolo, dar, din ce am spus până acum, putem să concluzionăm că însăşi ipseitatea unui sistem dinamic complex este

constituită de zona sa. Asemenea sisteme cu autoorganizare apar separându-se de orice altceva. Apoi, nu pot decât să îşi consemneze propriile stări: lumea nu-sistem nu poate fi „cunoscută” decât indirect, prin intermediul stărilor senzoriale ale păturii sistemului. Eu consider că aceste proprietăţi ale autoorganizării sunt de fapt precondiţiile esenţiale pentru subiectivitate.

Se impune o distincţie. Tocmai caracteristica de autoconservare a unor astfel de sisteme, tendinţa de a se separa de mediile lor şi apoi de a-şi menţine activ existenţa, constituie baza elementară a *ipseităţii.* Şi tocmai natura izolată a unor astfel de sisteme, faptul că pot să consemneze numai lumea nu- individuală prin intermediul stărilor senzoriale ale propriilor zone, constituie baza elementară a *subiectivităţii* – „punctul de vedere” al unui sine sechestrat.

Desigur, asta nu înseamnă că fiecare sistem cu autoorganizare are o subiectivitate *capabilă de a simţi.* Încă suntem departe de a putea identifica proprietăţile specifice pe care un sistem cu autoorganizare trebuie să le manifeste înainte de a putea deveni conştient. Totuşi, chiar dacă nu putem vorbi de conştiinţă încă, se pare că am dat peste un prototip fizic pentru problema altor minţi. Însăşi natura unei zone Markov constă în a determina o împărţire a stărilor în „sistem” şi „nu-sistem”, astfel încât stările nu-sistem sunt *ascunse* de interiorul sistemului şi invers[38].

Să revenim la supa primordială a lui Friston, unde lucrurile devin şi mai ciudate. După ce a generat în mod spontan un sistem cu autoorganizare dinamic şi

complex, Friston a testat dacă acest ansamblu permite *predicţii* cu privire la stările externe pornind de la stările interne ale sistemului. Dacă permite, a afirmat Friston, asta ar sugera că stările interne ale unui sistem îşi *modelează,* în timp, stările externe şi s-ar putea spune şi că *reprezintă* în interiorul lor acele evenimente externe. Ştiu că pare magie, dar asta înseamnă pur şi simplu că sistemul se adaptează la tiparele evenimentelor externe. (Ca să simplific mult ideea: aşa se explică faptul că poţi prezice direcţia obişnuită a vântului dintr-o zonă judecând după înclinaţia copacilor de acolo atunci când nu bate vântul. Înclinaţia copacilor „reprezintă" direcţia obişnuită a vântului, deoarece copacii au crescut sub acel unghi ca să se adapteze la vânt.)

Friston a examinat statutul funcţional al subsistemelor interne ale organismului pe care l-a simulat şi a descoperit tocmai această capacitate predictivă:

> Dinamica internă care prezice [un eveniment extern] pare să apară în fluctuaţiile sale înainte de evenimentul în sine – după cum s-ar anticipa dacă evenimentele interne modelează evenimentele externe. E interesant că subsistemul cel mai bine prezis a fost cel aflat la cea mai mare depărtare de stările interne. Acest exemplu ilustrează că stările interne deduc sau consemnează evenimente îndepărtate într-un mod ce nu e cu totul diferit de perceperea unor evenimente auditive prin undele sonore – sau modul în care peştii simt mişcarea în mediul lor. Subsistemele a căror mişcare a putut fi

prezisă cu un grad mare de precizie sunt cele mai semnificative de la periferia ansamblului, unde ansamblul are cea mai mare libertate de mişcare. Aceste mişcări sunt cuplate la stările interne – prin zona Markov – prin sincronism generalizat[39].

După ce a observat acest sincronism, prin care stările interne ale sistemului au modelat evenimente îndepărtate sub aspect fizic, Friston a concluzionat că stările interne manifestă „inferenţă”.

Aceasta se dovedeşte a fi cea mai semnificativă proprietate a unor asemenea sisteme. Zona Markov conferă stărilor interne ale sistemelor cu autoorganizare o capacitate de a reprezenta sub aspect probabilistic stări externe ascunse, astfel încât sistemul poate *să infereze cauzele ascunse ale propriilor stări senzoriale,* ceea ce seamănă cumva cu funcţia percepţiei. La rândul său, această capacitate permite sistemului *să acţioneze în mod intenţionat* asupra mediului exterior, în baza stărilor sale interne – acţiuni ce sunt similare activităţii motoare.

În acest fel, sistemul se menţine şi se reînnoieşte în fata perturbaţiilor externe[40]. Simplul fapt că este un sistem cu autoorganizare este suficient pentru a le conferi un scop lui şi fiecăreia dintre părţile sale, iar aceasta este funcţia stărilor active ale zonei: ele manipulează mediul pentru a menţine integritatea sistemului. Ceea ce înseamnă că, alături de un sine închis, un punct de vedere subiectiv, un obiectiv şi capacitatea şi de a simţi, şi de a acţiona, simplul fapt că există o zonă Markov aduce cu sine ceva similar unei

intenţionalităţi[41]*.* Poate că nu pare deosebit de impresionant în forma în care se vede în simulările lui Friston – punctele albastru-deschis (exterioare) le influenţează pe cele albastru-închis (interioare) prin cele violet (senzoriale), iar punctele albastru-închis se conectează iar la cele albastru-deschis prin cele roşii (active). Cu toate acestea, sper că înţelegi de ce sistemele biologice cu autoorganizare trebuie să infereze cauzele ascunse ale stărilor lor senzoriale – chiar dacă deocamdată în mod nonconştient. Dacă nu ar face-o, ar înceta să existe. Ele nu au cum să nu modeleze dependenţe cauzale din lume, astfel încât acţiunile lor din lumea respectivă să le asigure supravieţuirea.

De aici vine conceptul de „stări preconizate” şi din acest motiv sistemele biologice cu autoorganizare sunt homeostatice. Homeostaza pare să fi apărut odată cu autoorganizarea. Stările senzoriale şi active ale unei zone Markov nu sunt decât receptori şi efectori ai unui sistem cu autoorganizare, iar modelul stărilor externe pe care îl generează este centrul său de control.

Sistemele biologice cu autoorganizare *trebuie* să-şi testeze modelele lumii şi, dacă lumea nu oferă răspunsurile pe care le aşteaptă, ele trebuie să facă urgent ceva diferit, ca să nu moară. Prin urmare, deviaţiile de la stările preconizate sunt o formă elementară a „răspunsurilor evocate de echipament” ale lui Wheeler. Aşa apare formularea întrebărilor; *autoorganizarea îi face pe observatorii participanţi să existe.* Întrebarea pe care şi-o pune întotdeauna un

sistem cu autoorganizare este, pur şi simplu: „Supravieţuiesc dacă fac asta?”. Cu cât răspunsul e mai nesigur, cu atât mai rău pentru sistem.

Relaţia dintre stările active ale unei zone Markov şi autoconservare, prin homeostază, este ilustrată cel mai bine dacă arătăm ce s-a întâmplat când Friston *a cauzat leziuni* zonei sistemelor cu autoorganizare în simulările lui experimentale. A făcut acest lucru împiedicând selectiv stările senzoriale ale zonei să-i influenţeze stările active (vezi figura 14, graficele b, c şi d). În lipsa stărilor active obişnuite ale zonei Markov, a apărut haos entropic, iar sistemul s-a disipat rapid. Adică a încetat să existe.

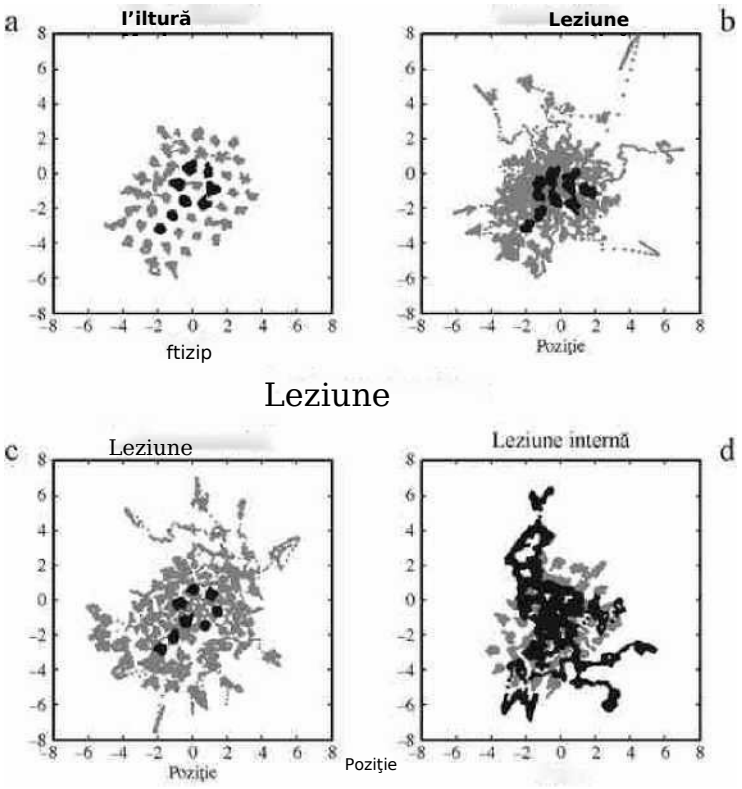


Figura 14. Efecte entropice ale unor leziuni uşoare la nivelul
zonei Markov a unui sistem cu autoorganizare

Ceea ce tocmai am ilustrat în termeni formali (folosind concepte din fizica statistică) este structura de bază a oricărui sistem cu autoorganizare dinamic. Friston rezumă relaţiile legitime după cum urmează: *„Orice sistem dinamic ergodic aleatoriu ce are o zonă Markov va părea că îşi menţine în mod activ integritatea structurală şi dinamică”*[42]. (Un sistem „ergodic” ocupă stări limitate.) Acest tip de activitate îndeplineşte criteriile lui Kant menţionate mai sus: o zonă Markov este atât scopul, cât şi mijlocul prin care un sistem cu autoorganizare persistă în timp – iar acest

lucru se întâmplă în mod natural. Astfel, asemenea sisteme par să aibă o minte proprie, chiar dacă foarte primitivă (şi nonconştientă).

Dacă această afirmaţie este corectă, atunci ar trebui să fie posibil să identificăm apariţia autoorganizării în orice ansamblu arbitrar de subsisteme cuplate cu interacţiuni pe distanţă mică. Exact asta a arătat experimentul cu supa primordială al lui Friston. Din acest experiment, el a extras patru proprietăţi fundamentale ale *tuturor* sistemelor biologice cu autoorganizare:
1) sunt ergodice;
2) au o zonă Markov;
3) manifestă inferenţă activă;
4) se autoconservă.

Ce anume este principiul energiei libere ce dă titlul acestui capitol? În primul rând, se ştie că este dificil de explicat. Iar asta în mare parte din cauza ecuaţiilor lui Friston, care, după cum am descoperit când am reuşit să înţeleg „Life aş we know it”, sunt şi greu de priceput, şi opace. Aşa că voi încerca să explic principiul numai cu ajutorul cuvintelor. Ecuaţiile pot fi întotdeauna exprimate în cuvinte, deoarece nu sunt decât afirmaţii despre nişte relaţii. Acestea fiind spuse, întrucât ecuaţia de bază îţi explică scopul şi obiectivele tale principale în viaţă, precum şi pe cele a orice altceva care a trăit vreodată, ar trebui probabil să o vezi în forma ei consacrată măcar o dată. lat-o:

$$A = U - TS$$

unde A înseamnă energia liberă, U este energia internă totală, T temperatura, iar S entropia.

Ce înseamnă asta? În termodinamică, energia liberă a unui sistem este egală cu cantitatea totală de energie conţinută în sistem minus partea folosită deja într-o activitate efectivă, deci care nu este liberă[43]. Aşadar, ecuaţia spune doar: „Energia liberă este egală cu energia internă totală minus energia utilizată deja”. Ce ar putea fi mai simplu? Energia liberă este ceea ce rămâne după ce îndepărtezi energia care nu este liberă (adică atunci când îndepărtezi „energia latentă”)[44].

Ecuaţia pe care tocmai am descris-o cuantifică cu acurateţe energia liberă în contexte termodinamice elementare, dar nu şi în contexte chimice, unde trebuie admisă o toleranţă pentru moleculele suplimentare ce sunt formate de unele procese la diverse temperaturi şi presiuni. De aceea, chimiştii folosesc o versiune uşor diferită a ecuaţiei[45]. Pentru a le deosebi, tipul termodinamic clasic de energie liberă este numită „energie liberă Helmholtz” (de la Hermann von Helmholtz, un membru important al Societăţii de Fizică din Berlin), iar tipul ansamblului chimic este numită „energie liberă Gibbs”.

Friston foloseşte o a treia versiune a aceleiaşi ecuaţii, pentru a cuantifica energia liberă în contexte *informaţionale.* Voi numi „energie liberă Friston” acest tip de energie liberă. Ecuaţia relevantă spune că „energia liberă Friston este egală cu energia medie minus entropia”[46]. Aici, „energie medie” înseamnă probabilitatea *preconizată* ca un eveniment să se întâmple conform unui model, iar „entropie” înseamnă incidenţa *reală* a concretizării modelului. Aşadar, energia liberă Friston este diferenţa dintre cantitatea de informaţii pe care te aştepţi s-o obţii dintr-un

eşantion de date – dintr-o secvenţă de *r r r* evenimente – şi cantitatea de informaţii pe care chiar le obţii. (Să ne amintim că entropia unui sistem predictiv este informaţia sa medie, iar creşterea informaţiei înseamnă descreşterea probabilităţii[47].) Prin urmare, ecuaţia „energia liberă Friston este egală cu energia medie minus entropia” spune, practic, acelaşi lucru ca ecuaţia „energia liberă Helmholz este egală cu energia internă totală minus energia care nu este disponibilă să acţioneze”[48]. Asta se întâmplă deoarece energia liberă Friston este analoagă cu energia liberă Helmholtz, unde există un schimb de *informaţii,* spre deosebire de un schimb *termodinamic* între sistem şi mediul său[49].

Dacă sistemele biologice trebuie să-şi minimizeze entropia, iar entropia este informaţie medie, atunci ele trebuie să menţină fluxul de informaţii pe care le prelucrează la un nivel minim. Trebuie să minimizeze evenimentele nepreconizate. Acest lucru este cunoscut sub denumirea tehnică de „surprindere”. La fel ca entropia, surprinderea este o funcţie descendentă a probabilităţii: pe măsură ce probabilitatea scade, surprinderea creşte[50]. Surprinderea măsoară cât de improbabil a fost un eveniment; entropia măsoară cât de improbabil se preconizează că va fi (în medie)[51]. Aşadar, surprinderea, ca şi entropia din model, este un lucru rău pentru organismele vii. Nu vreau să ne pierdem în detalii tehnice[52]. Vreau să spun doar că, la nivelul cel mai simplu, eşti într-o stare de surprindere dacă ieşi în afara setului de stări în care *te aştepţi* să te afli din punct de vedere biologic (de exemplu, sub sau peste 36,5-37,5°C sau respirând sub apă), tocmai

pentru că există o probabilitate mică să fii în acea stare.

Sistemele cu autoorganizare trebuie să minimizeze fluxul informaţional, deoarece creşterea cererii de informaţii presupune creşterea incertitudinii în modelul predictiv. Incertitudinea aduce surprize, care sunt rele pentru noi, ca sisteme biologice, întrucât pot fi periculoase. Dar cum putem să minimizăm surprizele prin minimizarea fluxului informaţional? Oare asta nu înseamnă să ne prefacem că realitatea nu există? Răspunsul este nu. Energia liberă Friston este un atribut cuantificabil al diferenţei dintre *modul în care lumea este modelată* de un sistem şi *modul în care lumea se comportă de fapt.* Aşadar, trebuie să minimizăm această diferenţă. Modelul de lume al unui sistem trebuie să se potrivească cât mai mult cu lumea reală, ceea ce înseamnă că trebuie să minimizeze diferenţa dintre datele senzoriale pe care le prelevează din lume şi datele senzoriale ce au fost prognozate de modelul său. Asta va maximiza „informarea reciprocă” dintre lume şi model, ceea ce minimizează incertitudinea.

O cale de a face acest lucru constă în *îmbunătăţirea modelului de lume al sistemului.* Erorile de predicţie pot fi transmise înapoi modelului generativ, astfel încât acesta să genereze predicţii mai bune data următoare. Mai puţine erori făcute înseamnă mai puţine erori de transmis modelului, ceea ce înseamnă mai puţin flux informaţional. Prin urmare, „informarea reciprocă” este un produs al comunicării, al formulării şi răspunderii la întrebări: se comportă lumea cum am prevăzut, da sau nu?

Dar, întrucât sistemele biologice ca mine şi ca tine sunt izolate de lume prin zona Markov a fiecăruia, nu ne putem compara direct modelele cu felul în care este lumea cu adevărat. De aceea, trebuie să ne aducem în cap întregul proces de minimizare a surprinderii, devenind atât „sursa”, cât şi „receptorul” informaţiei ce vine din întrebările pe care le-am pus.

Facem acest lucru prin măsurarea entropiilor relative – prin cuantificarea decalajului dintre stările senzoriale prezise de o acţiune şi stările senzoriale ce apar efectiv din acţiunea respectivă. De aici apare cantitatea numită energie liberă Friston, care este întotdeauna o valoare pozitivă mai mare decât surprinderea propriu-zisă.

Dovezile senzoriale (primite sub forma aşa-numitelor *spiketrains* generate în capul nostru – miliarde de unu şi de zero) sunt singurele date pe care le putem obţine. Din aceste date trebuie *să inferăm* structura cauzală a lumii. Pentru că suntem fiinţe cu o zonă Markov, suntem obligaţi să ne bazăm pe chestiuni precum distribuţiile de probabilitate, şi nu pe adevăruri absolute. Iată motivul pentru care este util să ştim că energia liberă Friston este întotdeauna mai mare decât surprinderea – permite creierului uman să aproximeze adevăruri necognoscibile folosind calcule statistice.

După cum am văzut în experimentul lui Friston cu supa primordială, modelele generative apar cu sisteme cu autoorganizare. Din acest motiv, ele sunt numite uneori sisteme „care se autovalidează”, deoarece modelează lumea în funcţie de viabilitatea lor, apoi caută dovezi pentru modelele lor. E ca şi cum ar spune

nu „Gândesc, deci exist”, ci „Exist, deci modelul meu este viabil”[53]. Modelul propriu fiecărui sistem biologic este determinat parţial de *specia* sa, cum am explicat deja. Tu – ca fiinţă umană – te aştepţi să ajungi în stări foarte diferite de cele ocupate în mod normal de un rechin. E foarte improbabil să ajungi să respiri apă la sute de metri adâncime în ocean, dar deloc improbabil pentru un rechin. Deci un anumit set de stări senzoriale este mai mult sau mai puţin surprinzător – mai mult sau mai puţin improbabil – în funcţie de specia organismului care se află în ele.

Testul unui model bun al sinelui-în-lume constă în a vedea în ce măsură permite sistemului autogenerat să interacţioneze cu lumea în moduri care îl menţin în limitele sale viabile. Cu cât aceste interacţiuni sunt mai bune, cu atât mai mică va fi energia sa liberă. Cu cât este mai mică energia sa liberă, cu atât mai multă din energia sistemului este transformată în lucru mecanic util, pentru autoconservare[54]. Astfel, principiul energiei libere explică în termeni matematici modul în care sistemele vii ca mine şi ca tine se opun celei de-a doua legi a termodinamicii printr-un lucru mecanic care menţine homeostaza.

De asemenea, explică alt mod în care sistemele cu autoorganizare se autovalidează (aproape că aş vrea să spun că sunt introspective): ele sunt obligate să-şi pună întrebări *lor înseşi* despre propriile stări. Mai exact, sunt obligate să întrebe: „Ce se întâmplă cu energia mea liberă dacă fac asta?”. Răspunsul la această întrebare va determina întotdeauna ce urmează să facă sistemul în continuare, într-un interval adecvat[55]. Acesta

este mecanismul cauzal din spatele tuturor comportamentelor voluntare[56].

În mod surprinzător, nu am ajuns încă la capătul puterilor psihice remarcabile ce se manifestă în supa primordială a lui Friston. Am văzut că până şi cele mai simple sisteme cu autoorganizare au individualităţi nonconştiente şi subiectivitate, intenţionalitate şi un scop (şi anume supravieţuirea). Ele percep şi, dacă este necesar, acţionează. La această listă impresionantă putem adăuga încă un trofeu. Se dovedeşte că ele manifestă un fel de raţionalitate. Putem să spunem chiar că sunt bayesiene.

Thomas Bayes a fost un preot şi teolog englez al cărui nume încă nu a fost uitat exclusiv datorită ideilor sale despre probabilitate, pe care nu s-a ostenit niciodată să le publice. Bayes ne-a învăţat (într-o lucrare publicată postum, în 1763) că trebuie să folosim *dovezi curente* împreună cu *cunoştinţe de fond* ca să facem şi să revizuim *cele mai bune conjecturi* despre lume. Cu alte cuvinte, folosind terminologia de-acum familiară a lui Friston, trebuie să prelevăm mostre senzoriale, să le comparăm cu predicţiile derivate din modelele noastre generative şi să ne actualizăm credinţele în mod corespunzător.

Iată formula standard[57]:

$$\mathbf{P(A|B)} \quad \frac{P(A|B)P(A)}{P(B)}$$

Exprimată în cuvinte, această teoremă spune: „Raportul probabilităţilor pentru două ipoteze condiţionate de un număr de date este egal cu raportul probabilităţilor lor condiţionale înmulţit cu gradul în care prima ipoteză o depăşeşte pe a doua ca predictor al datelor”.

Mai simplu: dată fiind o ipoteză urmată de nişte dovezi, trebuie să revizuieşti probabilitatea ipotezei analizându-i *verosimilitatea* coroborată cu *probabilitatea* sa *anterioară.* „Verosimilitatea” ipotezei este gradul de potrivire dintre ce prezice şi dovezile obţinute efectiv, iar „probabilitatea anterioară” constă în cunoaşterea ta de fond despre ipoteză (adică probabilitatea ei chiar înainte de a lua în considerare dovezile noi). Rezultatul este *probabilitatea posterioară* a ipotezei. Atunci, date fiind două convingeri concurente, conjectura ta cea mai bună este cea care

are cea mai mare probabilitate posterioară. (Vezi figura 16 de la p. 207.)

De exemplu: eşti la aeroportul din Cape Town şi vezi pe cineva venind dintr-o cursă de la Johannesburg care seamănă cu prietena ta Teresa. Ipoteza ta e că aceasta este Teresa. „Verosimilitatea” ei este probabilitatea că vezi pe cineva care arată aşa, dat fiind că ea este Teresa. Apoi îţi aminteşti că Teresa locuieşte la Londra, ceea ce reduce „probabilitatea anterioară” ca aceasta să fi ea. Concluzia ta („probabilitatea posterioară”) este că te uiţi la altcineva care doar seamănă cu Teresa[58].

Cel mai important aspect al teoremei lui Bayes pentru scopurile neuroştiinţelor este că explică modul în care funcţionează în realitate inferenţa perceptivă – un proces inconştient – şi modul în care funcţionează în realitate transmiterea semnalelor în procesarea senzorimotoare. Circuitele cerebrale calculează *concret* distribuţiile probabilităţii anterioare, apoi trimit mesajele predictive unor neuroni senzoriali, într-un efort continuu de atenuare a semnalelor de intrare; iar percepţia implică *concret* comparaţii între distribuţiile prezise şi cele reale, rezultând astfel calcule ale probabilităţii posterioare. Inferenţele rezultate constituie ceea ce *este, concret,* percepţia[59]. Percepţia este un efort de a autogenera semnalele senzoriale de intrare şi, prin aceasta, de a le suprima. Acesta este motivul pentru care atât de mulţi specialişti în neuroştiinţe vorbesc astăzi de „creierul bayesian”.

Să ne amintim ce am afirmat mai sus despre relaţia dintre „informaţii” şi lumea materială: deşi informaţiile nu pot fi văzute şi atinse concret, nu există niciun dubiu

în privinţa existenţei lor. Comportamentul sistemelor fizice este *determinat* de fluxurile informaţionale. De aici, minimizarea energiei libere Friston minimizează în acelaşi timp energia liberă Gibbs şi energia liberă Helmholtz, iar asta pentru că minimizarea erorii de predicţie minimizează fluxul informaţional, iar reducerea fluxului informaţional reduce *consumul metabolic* din partea creierului şi a corpului în ansamblu[60]. Acest lucru nu se întâmplă numai pentru că activitatea cerebrală consumă atât de multă energie (20% din total), ci şi pentru că minimizarea energiei libere statistice din creier reglează schimburile energetice fiziologice dintre corp şi lume[61]. Aceste schimburi presupun de asemenea consum metabolic. Astfel, se vede că creierul predictiv este „leneş” (pe termen lung): atent la orice ocazie de a obţine mai mult cu mai puţin efort[62].

Asta este o explicaţie minimalistă a ceea ce face creierul. Totuşi, ca să menţină organismul viu, creierul trebuie să facă mai mult decât să conserve resursele de energie – trebuie să ţină seama şi de numeroasele nevoi biologice (pe lângă echilibrul energetic) despre care am discutat în capitolul 5, aproape toate obligându-ne să efectuăm lucru mecanic în lumea exterioară. Vom vedea că diversitatea nevoilor ce ne caracterizează pe noi, ca organisme complexe, nu poate fi separată de conştiinţă.

Am aflat că suprimarea erorii de predicţie este mecanismul esenţial al homeostazei. Prin urmare, minimizarea energiei libere devine sarcina de bază a tuturor sistemelor homeostatice. Am văzut că, de fapt, ecuaţia energiei libere elaborată de Friston este o

reformulare, în termeni cuantificabili, a definiţiei date de Freud „pulsiunii” – „o măsură a cererii de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce e corporal” –, măsură considerată de Freud necuantificabilă. Acum o putem cuantifica. Forţa motrice fundamentală din spatele comportamentului voliţional al tuturor formelor de viaţă este că ele sunt obligate să-şi minimizeze energia liberă. Acest principiu guvernează tot ce
fac.

În formularea succintă a lui Friston, principiul energiei libere impune că toate cantităţile care se *pot* schimba – cu alte cuvinte, care fac parte din sistem – se vor schimba *cu siguranţă* pentru a minimiza energia liberă[63]. De fapt, această formulare e mai mult decât succintă, este o lege. Propun să o numim „legea lui Friston”: toate cantităţile dintr-un sistem cu autoorganizare ce se pot schimba se vor schimba ca să minimizeze energia liberă. Aşadar, iat-o. Înarmaţi cu această cunoaştere, tot ce numim viaţă psihică devine încadrabil în matematică.

Suntem aproape pregătiţi să aflăm ce este conştiinţa, în termeni formali şi mecanicişti. Dar mai întâi, o fabulă[64].

O ierarhie predictivă

Eve Periaqueduct, inginer constructor, este angajată ca să prevină şi să repare fisurile dintr-un baraj municipal. Ea nu ştie că obiectivul mai profund al jobului ei este să asigure o sursă sigură de apă şi energie pentru un sat din apropiere şi ca acesta să nu fie luat de viituri. Dar nu are nevoie să ştie – unica ei sarcină e să minimizeze fisurile barajului. (Poate că îşi aminteşte din timpul facultăţii că ea minimizează entropia barajului. Dar nu are nevoie să-şi amintească asta, sarcina ei este foarte practică.)

Angajatorii ei îi asigură echipamentele necesare şi o mică echipă de lucrători. Moşteneşte şi un set de instrucţiuni elaborate de predecesorii ei care indică punctele cele mai slabe ale barajului, spunându-i ce să facă şi când. Ea şi echipa ei întreţin şi repară barajul cu sârguinţă, concentrându-se proactiv asupra punctelor slabe şi astupând orice fisuri spontane, pe măsură ce apar. Treptat, de-a lungul anilor, îşi dă seama că şi unele dintre fisurile neprevăzute apar după nişte tipare regulate. Prin urmare, ea *actualizează* instrucţiunile care i-au fost transmise, devenind mai pricepută în

anticiparea (şi, astfel, prevenirea) fisurilor. Acest lucru reduce costurile.

Întreprinzătoarea Eve îşi dă seama că tiparele pe termen lung ale fisurilor pe care le consemnează se corelează cu condiţiile climatice. Fără să-şi dea seama, consemnările ei *creează un model* al climei locale (adică vremea şi consemnările ei implică „informare reciprocă”). Fără să-şi propună, ea generează un model al unui aspect al lumii de dincolo de baraj. Există tipare ale vremii, iar ele corespund tiparelor fisurilor.

Pornind de la această descoperire, Eve angajează personal suplimentar ca să înfiinţeze un departament de meteorologie, pe care îl numeşte unitatea ei de „detecţie meteorologică”. Acest lucru generează un nivel nou în ierarhia echipei sale, aflat într- un loc diferit, justificat de aşteptarea că obţinerea unor prognoze meteo mai bune va duce la diminuarea costurilor cu reparaţiile pe termen lung.

Acest nivel nou face modelul ei predictiv să fie mai sensibil la contexte preconizate. Nu e nevoie ca angajaţii departamentului de detecţie meteorologică să ştie că treaba lor are legătură cu prevenirea fisurilor; ei se concentrează exclusiv asupra sarcinii de a anticipa schimbările vremii. Eve le furnizează o schemă a condiţiilor preconizate derivate din instrucţiunile moştenite iniţial şi actualizate de ea. De reţinut că aceste instrucţiuni nu vizează tiparele preconizate ale fisurilor, ci condiţiile climatice preconizate.

Deoarece Eve nu vrea să piardă mult timp verificându-şi mesajele, noului departament i se cere să-i transmită rapoarte de feedback numai dacă apar

devieri de la condiţiile preconizate. Le numeşte rapoarte de „eroare” şi le foloseşte ca să-şi actualizeze în continuare schema condiţiilor climatice preconizate, pe care o retrimite apoi la staţia meteo, ştiind că astfel va reduce volumul de muncă al angajaţilor – şi, în cele din urmă, şi al ei.

Toate acestea permit departamentului de meteorologie să se axeze eficient pe sarcina trasată de ea. Ca să-şi îndeplinească sarcinile, angajaţii acestuia instalează o serie de instrumente de măsurare a vremii în locuri diferite, unele aflate la mare distanţă de baraj. Barometrele, termometrele, pluviometrele şi altele asemenea sunt calibrate de echipă astfel încât să trimită semnale staţiei meteo numai atunci când parametrii pe care îi măsoară (presiunea atmosferică, temperatura, umiditatea etc.) deviază de la limitele preconizate. Aceste limite sunt stabilite conform condiţiilor climaterice anticipate. Şi acest lucru eficientizează costurile, întrucât lucrătorii angajaţi să citească instrumentele (care creează încă un nivel ierarhic) nu trebuie să ajungă decât la cele care transmit semnale „de eroare” staţiei meteo. Pentru că ţine evidenţa acestor semnale, staţia poate să ajusteze periodic limitele preconizate pentru fiecare instrument, astfel *automatizându-şi* şi mai mult procedurile (adică reducând frecvenţa semnalelor care impun trimiterea pe teren a unor angajaţi care citesc şi reglează instrumentele).

Unii dintre algoritmii ce rezultă devin destul de subtili, pe măsură ce echipa descoperă că fluctuaţiile parametrilor pe care îi măsoară nu sunt neapărat

stabile şi regulate, ci variază în funcţie de *context.* De exemplu: „Dacă presiunea atmosferică e scăzută, trebuie să creşti volumul precipitaţiilor preconizate, dar numai iarna”. Chiar şi aşa, nu e nevoie ca angajaţii care citesc şi calibrează instrumentele să aibă informaţii despre sarcina mai amplă a departamentului de meteorologie. Unica lor sarcină este să citească şi să calibreze instrumentele, conform instrucţiunilor actualizate pe care le primesc de la biroul meteo. Şi mai puţin relevant pentru jobul lor este faptul că rapoartele departamentului de meteorologie despre devierile de la condiţiile climaterice preconizate îi sunt trimise lui Eve, pentru ca ea să poată prezice cu un grad mai mare de precizie tiparele fisurilor şi, prin urmare, să ia măsuri mai eficiente de mentenanţă a barajului.

De asemenea, sătenii din apropiere folosesc aceste prognoze meteo în alte scopuri, care nu au nicio legătură cu barajul. Astfel, ei au o impresie eronată despre obiectivul real al departamentului – cred că există ca să-i ajute să-şi planifice activităţile sociale în aer liber.

Eve Periaqueduct observă, în timp, că tiparul fisurilor din baraj se corelează nu numai cu condiţiile climaterice, ci şi cu evenimentele seismice. Prin urmare, înfiinţează o a doua echipă specializată, pe care o numeşte departamentul de „detecţie a cutremurelor”. Acesta se axează exclusiv pe generarea unor modele ale mişcărilor tectonice şi prezicerea lor, aşa că acest al doilea departament instalează, calibrează, monitorizează şi ajustează continuu propriile echipamente „senzoriale” tehnice. De asemenea, ţine

evidenţe complexe – care, ca şi în cazul departamentului de meteorologie şi al lui Eve însăşi, permit noii echipe să automatizeze anumite aspecte ale activităţii sale şi să se axeze numai pe fluctuaţiile imprevizibile pe termen scurt. (Şi sătenii folosesc prognozele în beneficiul propriu, deşi nu aceasta a fost menirea prognozelor.)

Astfel, se conturează gradual o ierarhie predictivă complexă, cu mai multe departamente, fiecare cu subniveluri proprii, ce măsoară diferiţi parametri din lumea de dincolo de baraj. Fiecare nivel al ierarhiei urmează numai instrucţiunile predictive actualizate pe care le primeşte de la nivelul imediat superior şi raportează numai devierile de la stările preconizate ale parametrilor monitorizaţi la acel nivel. Din punctul de vedere al lui Eve, rapoartele mixte pe care le primeşte de la departamentele ei senzoriale se contextualizează reciproc. Din când în când, ea trebuie să hotărască ce raport să *prioritizeze.* Până la urmă, dispune de resurse limitate şi nu se poate ocupa de toate evenimentele posibile.

Eve se ocupă în continuare de mentenanţa barajului în baza programului ei pe termen lung, de la care deviază numai atunci când nu se potriveşte cu prognozele combinate pe care le primeşte de la departamentele senzoriale. La rândul lor, departamentele îi trimit lui Eve rapoarte de feedback numai atunci când eşantioanele de date pe care le culeg deviază de la propriile predicţii stabilite de multă vreme. Şi tot aşa, până la angajaţii care citesc şi reglează instrumentele de măsură.

De asemenea, întregul proces al transmiterii de mesaje şi al actualizării programului ce se tot repetă între nivelurile organizaţiei înfiinţate de Eve respectă regula lui Bayes: foloseşte dovezile curente (mostre senzoriale) împreună cu cunoaşterea de fond (ipoteze anterioare) ca să facă *şi să-şi revizuiască* cele mai bune conjecturi (ipoteze posterioare) despre lume.

Cu timpul, munca lui Eve devine repetitivă şi plictisitoare şi ea abia aşteaptă pensionarea. Îşi dă seama că îşi pune problema: „Înainte să plec, aş vrea să construiesc un baraj nou şi mai bun”. Aşa că sună la conducerea municipală şi întreabă: avem cumva un departament „reproductiv”?

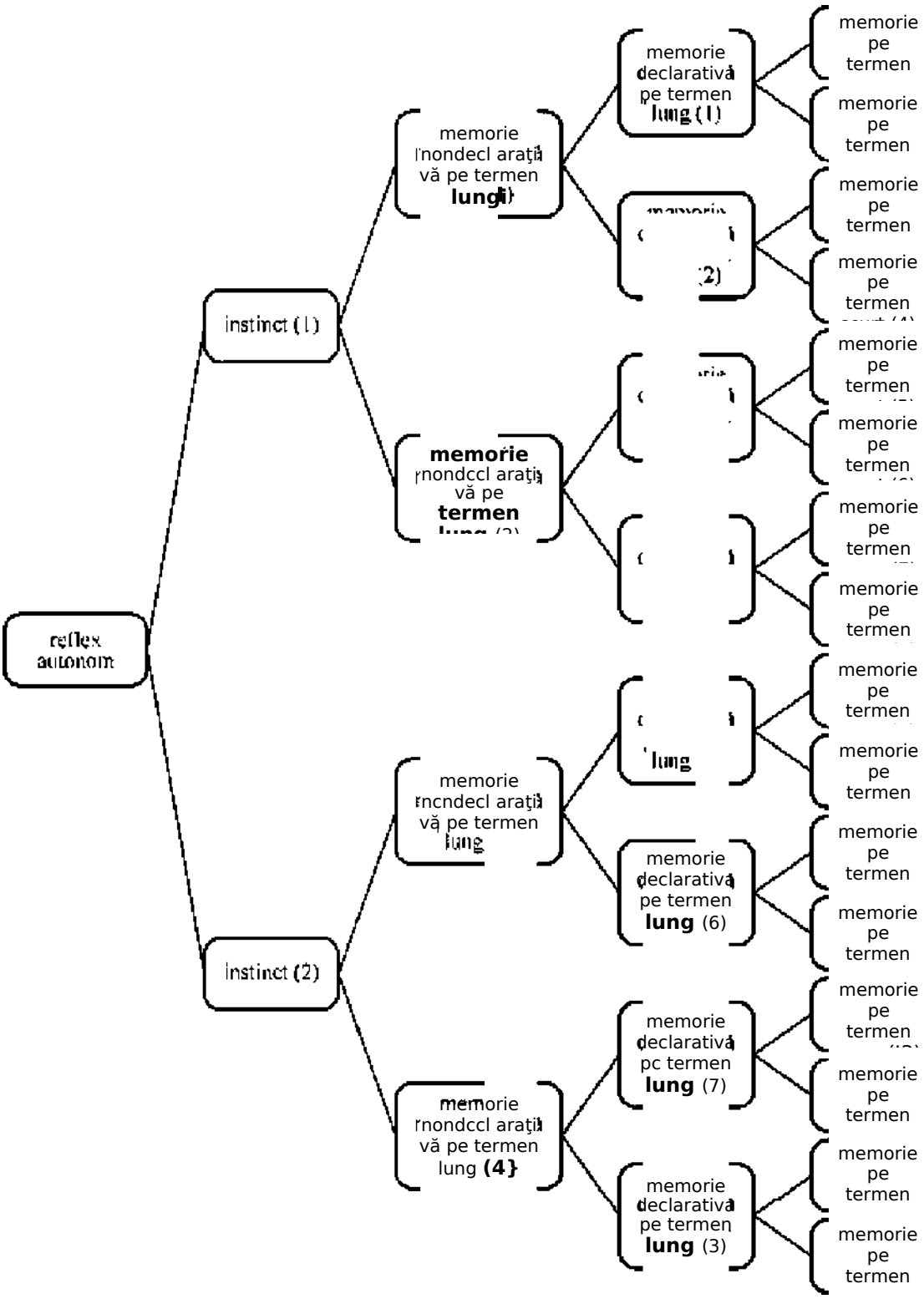
Foarte multe funcţii cvasipsihice apar la majoritatea celor mai simple sisteme cu autoorganizare, dar, ca să explicăm cum funcţionează un creier propriu-zis, trebuie să vedem cum se pot potrivi aceste sisteme într-o structură generală reciproc avantajoasă. Datorită studiilor lui Friston, este evident că sistemul nervos implementează o ierarhie predictivă iterativă care funcţionează în mare măsură ca şi cea stabilită de Eve Periaqueduct de-a lungul timpului. În cele din urmă, numeroasele funcţii complexe ale creierului chiar pot fi reduse la câteva asemenea mecanisme simple. După cum scrie Jakob Hohwy:

> Creierul încearcă oarecum cu disperare, dar cu pricepere să ţină sub control efectele pe termen scurt şi pe termen lung ale cauzelor ambientale asupra organismului, ca să-şi menţină integritatea.

Procedând astfel, apare implicit o reprezentare bogată, stratificată a lumii. Aceasta este o imagine frumoasă a minţii şi a locului nostru în natură, care ne îndeamnă la modestie[1].

De reţinut că în ştiinţele cognitive cuvântul „implicit” înseamnă *inconştient.*

În centrul modelului creierului despre sine în lume sunt generate predicţii specifice speciei despre limitele sale viabile (partea din stânga a figurii 15)2. Acestea se concretizează în reflexele autonome ce iau forma: „Dacă fac asta, temperatura mea va fi de aproximativ 37°C”. La nivelul următor (în dreapta), înconjurând acest centru, creierul generează comportamente instinctive (ce iau forma predicţiilor înnăscute pe care le-am descris în analiza emoţiilor de bază). La următorul nivel, generează comportamente involuntare dobândite (din sistemele sale de memorie nondeclarativă pe termen lung). La următorul, generează comportamente voluntare (din sistemele sale de memorie declarativă pe termen lung). Iar la final, la nivelul de la graniţa cu exteriorul, generează acţiunile de tip aici-şi-acum, cele mai nesigure care „prezic prezentul” (din sistemele sale de memorie pe termen scurt).



memorie declarativa pe termen **lung (**

memorie

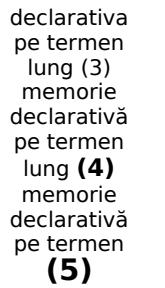


Figura 15. O ierarhie predictivă simplificată, extinzându-se de
la centrul autonom tonic la periferia senzorimotoare fazică.

Predicţiile circulă dinspre centru spre periferie (de la stânga la
dreapta, în această reprezentare); erorile de predicţie circulă în
direcţia opusă.

Simplific lucrurile, desigur: există mult mai multe

niveluri în ierarhia predictivă a creierului, nu doar cinci, aranjate în numeroase fluxuri de procesare paralele. Totuşi, se întrevăd nişte principii generale.

Primul este că creierul conspiră ca să anticipeze şi, astfel, „să elucideze” evenimentele din lume. Creierul suprimă semnalele de intrare predictibile, care nu au valoare de informaţie şi pe care altfel ar trebui să le prelucreze în mod inutil. Pe scurt, fiecare nivel din ierarhia sa primeşte numai informaţiile nepreconizate, demne de a fi ştiute, transmise de la nivelul imediat superior. Aceste rapoarte de feedback sunt erori de predicţie.

Al doilea principiu general este că această ierarhie se desfăşoară la scări temporale şi spaţiale progresiv mai mici. Predicţiile centrului se aplică în toate situaţiile, iar cele mai periferice sunt temporare şi focale. O secvenţă predictivă se desfăşoară de la nucleii care monitorizează corpul localizaţi în trunchiul cerebral şi în diencefal[3], prin ganglionii bazali şi sistemul limbic, prin neocortex, spre receptorii senzoriali specifici tipurilor de senzaţii localizaţi în organele terminale (de exemplu, în bastonaşele sau conurile retinei), care au câmpuri receptoare foarte înguste. La periferie, acurateţea pe termen scurt şi complexitatea prevalează în detrimentul generalizabilităţii pe termen lung, de care se bucură predicţiile mai profunde.

Al treilea principiu este conex: există o ierarhie a plasticităţii exprimată în predicţiile centrului care nu se pot schimba şi cele periferice care se pot schimba şi se schimbă; ele sunt supuse unei actualizări instantanee, nivelurile intermediare generând un grad intermediar

de plasticitate. Cu alte cuvinte, „centrul de control” al homeostatului creierului (modelul propriului sine) se actualizează constant – deşi manifestă o rezistenţă tot mai mare la schimbare, pe măsură ce valurile de erori se apropie de centrul său. Plasticitatea tot mai mare a nivelurilor sale mai periferice este unul dintre avantajele principale ale unui model predictiv *ierarhic.*

Al patrulea principiu, pe care nu l-am clarificat până acum, este că percepţia (spre deosebire de învăţare) *inversează direcţia* prelucrării informaţiilor. Prin inversarea dependenţelor cauzale care au influenţat iniţial modelul predictiv, creierul produce inferenţele noastre perceptive – pe care Merker le-a descris drept „o lume tridimensională, panoramică, bine definită, compusă din obiecte solide cu formă clară: lumea experienţei noastre fenomenologice familiare”[4]. (Aceste inferenţe circulă de la stânga spre dreapta în figura 15.) „Inversarea modelului predictiv” înseamnă pur şi simplu trecerea de la învăţare la predicţie în baza a ceea ce ai învăţat. E ceea ce făcea Eve Periaqueduct: infera („percepea”) starea lumii de dincolo de baraj din datele meteorologice şi seismologice pe care le primea.

Percepţia porneşte *din interior spre exterior,* întotdeauna din punctul de vedere al subiectului. De fapt e vorba de apercepţie, un proces deductiv, care ţine de testarea ipotezei în paradigma bayesiană[5]. Hermann von Helmholtz, primul care a înţeles esenţa acestei chestiuni, a numit-o „inferenţă inconştientă” (din nou, de notat adjectivul). Ceea ce vezi este „conjectura ta cea mai bună” cu privire la ce există în afara ta, e *propunerea ta de răspuns la întrebările pe*

care le adresezi acum lumii.

Creierul trebuie să infereze cauzele cele mai probabile ale semnalelor de intrare fără să aibă acces direct la lumea imposibil de cunoscut aflată dincolo de zona lui. Creierul nu poate să o ia decât în direcţia în care curg şi se modifică stările sale senzoriale *(spike trains;* figura 11). Sarcina lui este să folosească aceste semnale ca să creeze un *model* probabilistic al regularităţilor ce există în lumea reală (sau, mai degrabă, între el însuşi şi lume), pe care le foloseşte apoi ca să genereze inferenţe ce îi ghidează acţiunile – acţiuni care trebuie să-i asigure supravieţuirea în această lume. La rândul lor, acţiunile generează mostre senzoriale noi, ce sunt folosite pentru actualizarea în continuare a modelului, lucru pe care trebuie să-l facă pentru că modelele sunt imperfecte. Asta duce la alte acţiuni şi aşa mai departe.

Prin urmare, acţiunile trebuie privite ca *experimente* care testează ipoteze ce apar din modelul generativ. Dacă un experiment nu produce datele senzoriale preconizate, atunci sistemul fie (1) trebuie să-şi schimbe predicţia ca să explice datele mai bine, fie, dacă rămâne încrezător în predicţia iniţială, (2) trebuie să obţină date mai bune, adică să execute acţiuni care îi vor modifica inputul senzorial.

Aceste două opţiuni – schimbarea predicţiei sau a inputului – sunt mecanismele fundamentale ale percepţiei şi, respectiv, acţiunii.

Ultimele trei paragrafe de mai sus mă determină să corectez un bias ce influenţează acest capitol şi

capitolul anterior. Până acum, ca majoritatea specialiştilor care folosesc funcţiile corticale drept model al felului de funcţionare a creierului, mam axat aproape exclusiv pe inferenţa *perceptivă* bayesiană. Dar există şi inferenţă *activă.* De fapt, inferenţa activă este forma primară (cel puţin în biologie), întrucât percepţia există ca să ghideze acţiunea.

După cum tocmai am precizat, creierul bayesian are două moduri de a răspunde la eroarea de predicţie. Când se confruntă cu o ipoteză în cazul căreia se aplică o probabilitate „posterioară” în scădere, el creează o potrivire mai bună între ipoteză şi date schimbându-şi fie predicţia „anterioară”, fie inputul. Diferenţa dintre aceste alternative se reduce la *direcţia* statistică *a potrivirii:* eroarea este redusă dacă predicţia este schimbată astfel încât să se potrivească inputului senzorial şi este de asemenea redusă dacă inputul senzorial este schimbat astfel încât să se potrivească predicţiei. Bineînţeles, în realitate organismele pendulează tot timpul între aceste două opţiuni. (Să ne gândim la un şoarece-de-câmp care tot ţopăie prin tufăriş, oprindu-se ca să privească în jur, ţopăie din nou, se opreşte iar ca să privească şi tot aşa.) Din anumite puncte de vedere, percepţia şi acţiunea sunt mai asemănătoare decât par.

Corpul însuşi este o lume ascunsă „exterioară” zonei Markov a sistemului tău nervos central. Valurile de predicţii din figura 15 ar fi putut să includă straturi concentrice echivalente care să culmineze în receptorii şi efectorii viscerali terminali care determină funcţionarea organelor tale *interne.* Aşadar, modelul

despre lume al creierului trebuie să includă un model al sinelui tău corporal şi al traiectoriei sale în aceeaşi măsură în care trebuie să includă un model al tuturor celorlalte cauze ascunse ce prezintă interes pentru tine. (În exemplul meu de mai sus, nu ai fi putut ieşi din încăperea plină cu fum dacă nu ai fi avut un model implicit al modului în care se mişcă corpul tău *şi* în care îţi echilibrează gazele sangvine.)

Mai mult, acţiunea nu apare deoarece creierul transmite vreun mare plan tuturor muşchilor şi organelor corpului. În schimb, muşchii se contractă, iar glandele secretă până când dispar erorile de predicţie pe care le transmit în ierarhie. Astfel, organele „acţiunii" musculoscheletale şi viscerale ale corpului sunt la cheremul semnalelor de eroare generate prin diferenţele dintre ce se aşteaptă modelul predictiv să realizeze ele şi ce realizează ele de fapt. Suprimarea erorii de predicţie este ceea ce controlează acţiunea, nu mai puţin decât percepţia[6].

Să ne amintim de legea lui Friston: toate cantităţile dintr-un sistem cu autoorganizare ce se pot schimba se vor schimba ca să minimizeze energia liberă. Numeroasele homeostate corporale reglate şi orchestrate de metahomeostatul din mezencefal constituie pivotul mecanismului prin care ne menţinem în viaţă, din simplul motiv că reglarea homeostatică ne ţine corpul în limitele sale viabile. Aceste limite nu pot fi schimbate. Asta înseamnă – conform legii lui Friston – că *altceva din sistem trebuie să se schimbe.* Aceasta este explicaţia formală, mecanicistă a legăturii imperative dintre pulsiune şi *acţiune,* precum şi motivul

pentru care trebuie să existe *o ierarhie a predicţiilor anterioare,* dintre care unele pot fi schimbate, iar altele nu[7].

Totuşi, nu prea este de folos să acţionezi orbeşte. Acţiunea trebuie să fie ghidată de percepţie, care este generată de un model al sinelui în lume. Regula lui Bayes descrie felul în care este implementat modelul predictiv ce realizează acest lucru, modul în care se actualizează, astfel, constant, precum şi motivul pentru care trebuie actualizat. Aceasta devine baza formală a *învăţării,* prin care predicţiile sunt obţinute şi nuanţate în timp din semnalele de eroare primite. De asemenea, aceste dinamici ale sistemului confirmă, din raţiuni la fel de mecaniciste, că acţiunea are prioritate faţă de percepţie: numai acţiunea poate să crească probabilităţile predicţiilor anterioare – dintre care unele, după cum am spus, pur şi simplu nu pot fi schimbate.

Regula lui Bayes pleacă de la o asumpţie a ceea ce am numit „cunoaştere de fond”. Altfel, ea nu poate funcţiona[8]. Apare astfel o întrebare: de unde vine cunoaşterea de fond, la început, înainte ca sistemul să adune dovezi despre lume? Răspunsul este că principalele noastre „stări prognozate” sunt codificate de specia noastră ca puncte de stabilizare homeostatice *înnăscute* – cantităţi care au fost determinate de ceea ce a funcţionat efectiv pentru strămoşii noştri pe parcursul evoluţiei. Noi suntem beneficiarii succeselor biologice ale generaţiilor trecute, care stabilesc premisele elementare ale existenţei noastre.

Şi nu ne putem culca pe laurii lor. Legătura dintre

afect şi acţiune dictează că, dacă nu reuşeşti de la început, trebuie să tot încerci. Cererile celor mai profunde pulsiuni biologice ale tale sunt inexorabile: intensitatea lor scade numai dacă le satisfaci sau dacă mori. Cu excepţia celei de-a doua situaţii, trebuie să-ţi suplimentezi reflexele şi instinctele înnăscute şi să dezvolţi *alte* modalităţi de a-ţi satisface nevoile. Nu există alternativă. Cu alte cuvinte, trebuie să înveţi din experienţă. Din fericire pentru noi, creierul uman este neobişnuit de bine echipat pentru asta.

O implicaţie interesantă a tuturor acestor lucruri este că, dacă afectul chiar funcţionează conform mecanismelor homeostatice pe care le-am descris – dacă într-adevăr este „o măsură a cererii de travaliu impuse psihicului ca urmare a legăturii sale cu ceea ce e corporal” –, atunci el trebuie să fie *adevăratul* vehicul fundamental al minimizării energiei libere. Aşadar, afectul este mediul primar al voliţiunii – şi izvorul întregii vieţi psihice.

Am spus mai sus că învăţarea din experienţă produce un model ierarhic al lumii care, atunci când este inversat, generează predicţii despre aceeaşi lume. Dar procesul nu se încheie aici – predicţiile trebuie *testate.* De aici apar erori de predicţie, care sunt folosite pentru actualizarea modelului. Asta înseamnă învăţare „din experienţă”. Erorile de predicţie sunt semnalele senzoriale care *nu* au fost prezise de o ipoteză curentă, deci cele care nu au fost autogenerate. Aceasta este partea esenţială a datelor.

Greşeala pe care o fac în acest punct majoritatea specialiştilor în ştiinţe cognitive este să presupună că

datele de intrare sunt exclusiv *exteroceptive.* Ei uită că erorile de predicţie (inputurile senzoriale) care contează cel mai mult pentru noi vin din *interior*. De exemplu, devierea de la temperatura corporală preconizată furnizează feedback „senzorial”, nu mai puţin decât evenimentele exterioare nepreconizate; la fel şi semnalul de eroare homeostatică ce declanşează alarma de sufocare. Aceste semnale generează *afecte,* şi nu percepţii. După cum a afirmat Freud, prozencefalul este un „ganglion simpatic”. Confuzia din acest punct este preţul etern pe care îl plătesc colegii mei pentru că acceptă eroarea privitoare la cortex[9]. Conştiinţa este generată la nivel endogen – toată. Conştiinţa, la origine, este afect. Apoi este extinsă spre exterior către percepţie, ca să evalueze inferenţele perceptive, în modul pe care îl voi descrie în continuare.

În sfârşit, ne putem ocupa de întrebarea: de ce şi cum anume *devin conştiente* funcţiile naturale de autoconservare descrise în acest capitol? Ştim că conştiinţa e înrădăcinată în afecte, în sentimente. Dar care sunt legile formale, mecaniciste care dau naştere sentimentelor – şi, astfel, conştiinţei?

De ce şi cum se naşte conştiinţa

j III

Întrebarea fundamentală pe care trebuie să şi-o pună mereu fiinţele vii este: „Ce se întâmplă cu energia mea liberă dacă fac asta?”. Dar să faci ce? Direcţiile potenţiale de acţiune în orice moment nu sunt arbitrare sau infinite, ci sunt dictate de *nevoile* curente. Există, desigur, o legătură strânsă între nevoi şi acţiuni – fiecare nevoie cere o anumită acţiune adecvată. Dacă îţi este foame, trebuie să mănânci. Dacă eşti obosit, trebuie să te odihneşti. Totuşi, există obstacole în calea execuţiei: nu poţi face decât unu sau două lucruri în acelaşi timp. Asta înseamnă că, pentru a-ţi selecta următoarea acţiune, trebuie să-ţi ordonezi nevoile curente în funcţie de urgenţă.

Există două lucruri care fac sarcina ordonării mai complicată decât ar putea să pară la prima vedere. Primul: de obicei, nu e necesar ca nevoile noastre, de organisme complexe, să fie satisfăcute într-o ordine fixă. Să mănânci sau să dormi – ce e mai important? Depinde de tot felul de consideraţii. Al doilea: multe dintre nevoile organismelor complexe nu pot fi satisfăcute

întotdeauna prin aceeaşi acţiune. Ca să mănânci supă, îţi trebuie abilităţi diferite de cele necesare ca să mănânci porumb fiert, de exemplu (şi nici nu mă gândesc la abilităţile şi resursele implicate în prepararea lor). În ambele cazuri, „ce faci în continuare” depinde neapărat de context. De aceea, nevoile interioare trebuie prioritizate în funcţie de *condiţiile exterioare predominante.*

Ne naştem având predicţii specifice speciei despre ce e de făcut în stări precum foamea, setea, frica şi furia. Aceste predicţii înnăscute sunt numite „reflexe” şi „instincte” – instrumente de supravieţuire moştenite pentru care avem toate motivele să fim recunoscători. Dar ele nu sunt suficient de flexibile ca să facă faţă diversităţii şi complexităţii situaţiilor cu care ne confruntăm de fapt, aşa că trebuie completate. Acesta este rolul învăţării din experienţă.

Am arătat deja legătura dintre învăţarea din experienţă şi legea afectului. Valenţa afectivă – sentimentele noastre despre ce este „bun” şi „rău” din punct de vedere biologic pentru noi – ne ghidează în situaţii *neprevăzute.* Am concluzionat că acest fel de a ne găsi drumul printre problemele neprevăzute ale vieţii în funcţie de ceea ce simţim, cu ajutorul comportamentului *voluntar,* este funcţia biologică a conştiinţei. Ea ne ghidează alegerile când suntem în întuneric. Dar, desigur, ca să poată face acest lucru, ea trebuie să creeze legătura dintre afectele noastre interioare (înrădăcinate în nevoi) şi reprezentările lumii exterioare.

Acest lucru explică de ce activarea este însoţită de

sentimente *şi* de percepţii conştiente ale lucrurilor. Am recunoscut la sfârşitul capitolului 6 că marele mister al acestei legături – misterul modului în care experienţa subiectivă se potriveşte în materia universului fizic – poate fi elucidat numai dacă reducem atât fenomenele fiziologice, cât şi cele psihologice la cauzele lor mecaniciste fundamentale. Aceste cauze urmau să fie dezvăluite la un nivel de abstracţie pe care numai *fizica* îl putea furniza. În ultimele două capitole am demarat o descriere formală a acestor mecanisme unificatoare.

Acum a sosit timpul să finalizez descrierea. Dacă autoorganizarea şi homeostaza nu explică în sine de ce şi cum apare conştiinţa, atunci care este explicaţia? Care este legătura, formal şi mecanic, dintre procesul biologic al *prioritizării nevoilor* pe care tocmai l-am rezumat şi minimizarea energiei libere? Şi cum se face că rezultatul acestui proces *este simţit ca ceva* de către şi pentru unele sisteme cu autoorganizare?

Punctul de plecare al răspunsului meu este exact ceea ce tocmai am evidenţiat: creaturile complexe ca noi, vertebratele, au multe tipuri de nevoi. Adică avem multe sisteme interne, fiecare reglat de propriile mecanisme homeostatice, toate contribuind cu valori de eroare la calculul total al energiei libere. Nevoile noastre biologice sunt aceste valori de eroare. Când nevoile sunt resimţite ca afecte, le descriem ca având „valenţe” pozitive sau negative. Asta înseamnă că au o *valoare* subiectivă: ni se par bune sau rele pentru noi. Behavioriştii au încercat să obiectiveze valoarea redefinind sentimentele plăcute şi neplăcute drept

stimuli care recompensează şi care pedepsesc – dar ne-am ocupat deja de această chestiune. Valenţa nu ţine de stimul, ci este, inerent, subiectivă şi calitativă. Ceva poate fi încântător pentru o persoană şi înfricoşător pentru alta.

Dar este oare *imposibil* să cuantificăm valenţa? Să analizăm figura 12: cu cât săgeata deviază spre dreapta, cu atât mai mare este neplăcerea. Astfel, în orice moment, valoarea foamei tale ar putea fi 3/10 (care e mai rea decât 1/10), iar valoarea setei tale ar putea fi 2/10 (care e mai bună decât 5/10) etc. Specialiştii în afecte fac tot timpul asemenea măsurători, le cer participanţilor la cercetare să-şi evalueze plăcerile şi neplăcerile pe ceea ce se numeşte scară Likert. Deşi aceste scări sunt subiective, faptul că sunt cuantificabile *în principiu* lasă deschis răspunsul la întrebarea: de ce trebuie calificate afectele? Dacă sistemele cu autoorganizare pot să consemneze „răspunsurile evocate de echipament” (adică propriile stări) drept cantităţi, în principiu, atunci care este contribuţia calităţii? Această întrebare se referă la ceea ce filosofii numesc „qualia” – materia minţii care ne scapă şi despre care se spune că nu i se poate găsi locul în concepţia noastră fizicalistă asupra universului.

Răspunsul începe de la faptul că nevoile nu pot fi combinate şi adunate într-un mod simplu. *Multiplele noastre nevoi nu pot fi reduse la un singur numitor comun,* ci trebuie să fie evaluate pe scări separate, aproximativ egale, astfel încât fiecare să fie recunoscută. Nu poţi spune pur şi simplu că „3/10 din foame plus 1/10 din sete este egal cu 4/20 din nevoia

totală”, apoi să încerci să minimizezi valoarea totală, deoarece fiecare nevoie în parte trebuie satisfăcută. Metabolismul energetic nu este acelaşi lucru cu hidratarea, care nu este acelaşi lucru cu termoreglarea şi aşa mai departe; fiecare este esenţială. Conform lui Edmund Rolls, specialist în neuroştiinţe şi behaviorist, „dacă recompensa alimentară ar fi întotdeauna mult mai puternică decât alte recompense, atunci genele animalului nu ar supravieţui, pentru că el nu ar bea niciodată apă”[1].

Dacă analizăm aceşti factori laolaltă, are sens ca sistemele biologice cu autoorganizare să îşi diferenţieze nevoile (valorile de eroare) pe criterii *categoriale.* Deosebirea dintre variabilele categoriale este *calitativă.* Întrucât nu se poate pune semnul egalităţii între tipul de eroare *A* de 8/10 şi tipul de eroare *B* cu aceeaşi valoare, din motivele pe care tocmai le-am explicat, ele trebuie tratate ca variabile categoriale. Astfel, sistemul poate să o recunoască pe fiecare în parte pe termen lung *şi* să le prioritizeze în funcţie de context. De aceea, are sens ca sistemele complexe care se autovalidează să-şi clasifice numeroasele homeostate, astfel încât să le poată estima independent unul de celălalt şi să prioritizeze efectele.

Nu numai că nevoile diferite contribuie cu cantităţi diferite la energia liberă totală, ci şi cantităţile diferite au implicaţii diferite pentru animal în contexte diferite (de exemplu, foamea trece înaintea somnolenţei în unele situaţii, dar nu şi în altele). Acest lucru contribuie semnificativ la incertitudine – duşmanul de moarte al mecanismelor predicţiei. Incertitudinea tot mai mare

este o situaţie periculoasă pentru orice sistem cu autoorganizare: prezice moartea sistemului. Mai multă incertitudine cere mai multă complexitate de calcul (adică mai mult flux informaţional, adică mai multă entropie). Astfel, clasificarea devine o necesitate atunci când valoarea relativă a unor cantităţi diferite *se schimbă* în timp (dacă 8/10 pentru *A* valorează acum, dar nu întotdeauna mai mult decât 8/10 pentru *B*).

Ne putem imagina că s-ar putea dezvolta un set extrem de complex de algoritmi ai modelului pentru a calcula cereri relative de supravieţuire în toate situaţiile predictibile, pentru a ne permite *automat* să prioritizăm acţiunile pe acest criteriu. Totuşi, asemenea modele complexe sunt extrem de costisitoare, în toate sensurile cuvântului. Sunt greu de gestionat, ceea ce înseamnă întârziere, ce poate face diferenţa dintre viaţă şi moarte; necesită o mare capacitate de procesare, ceea ce înseamnă că trebuie să găsească mai multe resurse energetice. Statisticienii numesc „explozie combinatorie” creşterea exponenţială a resurselor de calcul ce sunt necesare unei creşteri liniare a complexităţii modelului.

Mai mult, un model complex care prezice cu acurateţe ce se întâmplă într-o situaţie specifică probabil că nu va prezice cu aceeaşi acurateţe ce se va întâmpla în alte situaţii. În termeni statistici, spunem că modelele excesiv de complexe „supraajustează” un eşantion de date. Modelul fisurilor preconizate în baraj alcătuit de Eve Periaqueduct nu se baza pe evenimente monitorizate oră de oră sau zi de zi în cele câteva săptămâni anterioare, ci pe nişte medii pe termen lung,

calculate pe baza unui număr mare de eşantioane de date culese de-a lungul mai multor ani. Acest lucru i-a făcut modelul mai simplu şi, de aceea, mai *generalizabil.* Conform „briciului lui Occam” (principiul parcimoniei)[2], vrem modele predictive *simple.* Simplificarea este esenţială pentru ca modelele noastre să se aplice într-o mare varietate de situaţii. Trebuie să fie utile, nu numai aici şi acum, ci şi în multe alte contexte.

Aşadar, modelele predictive trebuie să fie simple. Însă, după cum a spus Einstein, „totul trebuie făcut cât de simplu se poate, dar nu mai simplu”[3]. Cum ajungem la echilibrul corect? Compartimentarea este metoda statistică standard folosită pentru a obţine echilibrul optim între complexitate şi acurateţe. Asta se poate face în mai multe feluri. De exemplu, o parte din creierul vizual calculează *ceea ce* vezi, iar o alta calculează *unde* se află acel ceva. Astfel, ai posibilitatea să presupui identitatea constantă a ceva în timp ce se mişcă în jurul tău, îşi schimbă forma, mărimea şi orientarea. Dar cel mai mult contează că o capacitate de a compartimenta permite sistemului să-şi ordoneze nevoile şi predicţiile lor aferente (adică sursele predominante ale energiei libere preconizate) în categorii, de-a lungul timpului, şi să-şi concentreze eforturile de calcul asupra compartimentului prioritizat.

Aceasta este baza statistico-mecanică a realităţii că fiecare afect are nu numai o valenţă hedonică *continuă* (un *grad* de plăcere şi de neplăcere – aspect comun tuturor afectelor), ci şi o *calitate categorială* (astfel încât, de exemplu, setea se simte diferit de detresa de

separare, care se simte diferit de dezgust şi aşa mai departe). Acestea sunt caracteristici esenţiale ale qualia afective, forma elementară a tuturor qualia: ele au atât cantitate, *cât şi calitate.* Mai exact: afectele sunt întotdeauna subiective, calitative şi au valenţe. Ele trebuie să fie aşa, întrucât au evoluat astfel încât să rezolve problema controlului.

E util să descriem selectarea şi prioritizarea în funcţie de categorii ale afectelor în termeni de modelele operaţionale. Să ne gândim la felul cum se comportă un avion când este în modul „decolare” versus modul „croazieră” versus modul „turbulenţe” versus modul „aterizare”. În aceste situaţii diferite sunt implicate aceleaşi variabile, dar ele trebuie *cântărite* diferit de fiecare dată. De exemplu, altitudinea exactă contează mult mai mult în timpul aterizării decât în timpul regimului de zbor de croazieră. Acelaşi lucru se aplică în cazul barajului de care se ocupa Eve Periaqueduct. În modul „iarnă” versus modul „vară” şi în modul „cutremur” versus modul „noncutremur” se aplicau programe de funcţionare diferite. În condiţii de cutremur, Eve ar fi obligată să ignore programul sezonier normal, automatizat, ca să implementeze un program de tip
„urgenţă de cutremur”.

În termenii fiziologici pe care i-am folosit în capitolul 6, diferitele moduri de operare sunt funcţii *de stare* ale creierului. Aşa cum am explicat, triunghiul de decizie din mezencefal selectează stări cerebrale afective – precum modul „alarmă de sufocare”. Asta se întâmplă când PAG, substanţa cenuşie periapeductală, răspunde

la întrebarea: „Care dintre aceste semnale de eroare convergente (adică nevoi) constituie cea mai mare ocazie de a-mi minimiza energia liberă?”. Cu alte cuvinte, care nevoie este predominantă *în acest moment?* Răspunsul vine nu numai din amploarea relativă a semnalelor de eroare concurente, ci şi din diferenţele dintre categorii (moduri sau stări), a căror predominanţă trebuie evaluată în context. După cum am explicat, informaţiile contextuale sunt furnizate de coliculii superiori.

Iată un exemplu pe care l-am observat astăzi. Când am ieşit să fac jogging, la 7 dimineaţa, era întuneric şi, când m-am întors acasă, o oră mai târziu, era lumină. (E iarnă şi locuiesc într-o zonă rurală din Sussex, unde scriu această carte.) La plecare, am trecut pe lângă un câmp aflat în apropierea unei ferme, unde o turmă de oi m-a observat şi numai nu s-au călcat în picioare înghesuindu-se să fugă de mine. La întoarcere, când am trecut pe lângă acelaşi câmp, aceleaşi oi, aflate în acelaşi loc, abia dacă s-au uitat la mine. Spaima lor în contextul întunericului a fost înlocuită de plictiseală în timpul zilei. Pe scurt, contextul a modificat semnificaţia evenimentului „om care aleargă spre mine”. Noaptea, acest eveniment este prioritizat, fapt care introduce brusc oile în modul FRICĂ, dar nu şi ziua, aşa că oile rămân în modul implicit CĂUTARE.

Astfel se determină ce face sistemul în continuare. Cu alte cuvinte, se determină stările *active* care vor fi selectate de modelul generativ pentru a rezolva categoria de incertitudine ce a fost prioritizată. E ca şi cum sistemul spune: în condiţiile actuale, *aceasta este*

categoria de procesare a erorii de predicţie în care nu poate fi sacrificată complexitatea de calcul. Aşa că sistemul (în acest caz, oaia) trece în modul de operare FRICĂ. Aplică strategia cea mai bună pe care o poate furniza modelul său generativ în circumstanţele date: fuge. Apoi, după ce a luat în considerare tot ce a învăţat despre câmpul pe care se află (contextul preconizat), speră la ce e mai bine, dar se pregăteşte pentru ce e mai rău. Adică îşi execută planul cu prudenţă, gata să se adapteze la evoluţiile în desfăşurare.

Este esenţial că trecerea în modul FRICĂ înseamnă că nevoia prioritizată a devenit un afect. Cu alte cuvinte, a devenit conştientă. De ce? Devine conştientă pentru ca *devierile de la rezultatele preconizate* în categoria de nevoi predominantă să fie simţite în întreaga ierarhie predictivă. Iată ce este afectul. E „răspunsul evocat de echipament” la întrebarea despre sine pe care şi-a pus-o sistemul: „Care dintre aceste semnale de eroare convergente constituie cea mai mare ocazie de a-mi minimiza energia liberă?”.

În Africa de Sud, unde locuiesc în cea mai mare parte a timpului, ai nenumărate ocazii să vezi cum funcţionează în condiţii naturale selectarea afectului, adică în tipurile de condiţii în care a evoluat iniţial acest mecanism. Mă refer la ce se întâmplă nu între lei şi gazele în rezervaţiile noastre naturale fabuloase (cu toate că şi aici găseşti multe ocazii), ci în societatea noastră foarte inegală şi deci afectată de multe probleme. Mulţi dintre compatrioţii mei ştiu cum e când nevoia ta predominantă este să scapi de cineva care încearcă să te

omoare. În acest punct, comportamentul tău voluntar devine dominat de sentimente de FRICĂ, ce evaluează succesul sau eşecul aici-şi-acum al acţiunilor tale aflate în desfăşurare. Alte nevoi (precum nevoia de a urina) sunt pasate modului automat. Cu alte cuvinte, e posibil să faci pe tine fără ezitare.

Încercăm să reducem fenomenele fiziologice şi psihologice ale activării afective la un set de principii mecanice care pot fi formalizate matematic. În capitolul 6 am explicat că, odată ce triunghiul de decizie din mezencefal a prioritizat o nevoie, modelul prozencefalului despre sinele său în lume generează un context preconizat în care nevoia respectivă va fi satisfăcută. Am spus că această lume preconizată are două faţete: pe de o parte, ea reprezintă conţinutul real al predicţiilor noastre, iar pe de altă parte, trebuie să codifice nivelul nostru de încredere în predicţii. Prima dintre aceste faţete este furnizată de reţelele memoriei pe termen lung din prozencefal, care filtrează prezentul prin prisma trecutului. Am prezentat în capitolul 7 principiile care le guvernează. A doua faţetă – ajustarea nivelurilor de încredere – este modulată de „activare”. Aşadar, să formalizăm legile care o guvernează.

Primul mecanism identificat până acum este cel prin care este selectată categoria predominantă de acţiune (cea mai eficientă modalitate sau stare de funcţionare). Acesta este modul în care anumite calităţi afective mai întâi reglează acţiunile unor sisteme complexe cu autoorganizare. Asta duce la generarea unor planuri

senzorimotoare prin care sistemul „speră la ce e mai bine, dar se pregăteşte pentru ce e mai rău”. Iată detaliile asupra cărora vreau să zăbovesc acum. Care este mecanismul cauzal prin care se reglează speranţa la ce e mai bine şi pregătirea pentru ce e mai rău?

Prima parte a răspunsului este că nivelurile de încredere ataşate predicţiilor sunt *învăţate din experienţă,* la fel ca orice altceva. Apoi, nivelul de încredere în predicţiile noastre poate fi anticipat, ca şi predicţiile. Codificarea predictivă ne impune să asociem probabilităţi stărilor senzoriale care preconizăm că vor decurge din anumite acţiuni, iar apoi să comparăm probabilităţile cu distribuţiile observate concret în eşantioanele senzoriale ce urmează. Aceasta este esenţa metodei bayesiene de „actualizare a ipotezelor” pe care am descris-o mai sus – metoda prin care ne minimizăm energia liberă.

Acum, ca să stabilim potrivirea dintre un model şi unele date, nu e suficient să comparăm pur şi simplu *media* distribuţiilor, ci e necesar şi să evaluăm şi *variaţia mediei* (vezi figura 16). O valoare mare a variaţiei într-un eşantion reduce încrederea în potrivire. Dacă o ştire spune „A murit regele... A murit regele. A murit regele”, e mai probabil să o iau în serios decât dacă spune „A murit regele. nu, regele nu a murit. de fapt, se pare că regele s-ar putea să fi murit până la urmă”. Evaluările cu privire la diferenţa dintre o distribuţie prevăzută şi un eşantion de date sunt mai uşor de făcut atunci când distribuţia este îngustă şi *precisă.*

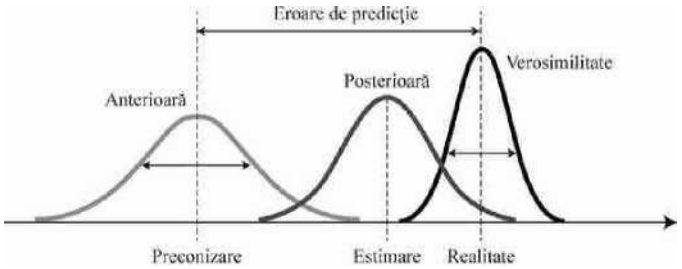


Figura 16. O predicţie anterioară (stânga) este comparată cu un eşantion de date senzoriale (dreapta), rezultând o predicţie posterioară (mijloc). „Mediile” acestor trei distribuţii sunt indicate de liniile verticale (punctate), iar „variaţia” lor este indicată de liniile orizontale (cu săgeţi). Variaţia preconizată mare în distribuţia anterioară (linia orizontală din stânga) indică un grad scăzut de încredere în predicţia anterioară, iar variaţia reală îngustă din eşantionul de date senzoriale (linia orizontală din dreapta) indică un grad ridicat de încredere în date. În acest exemplu, precizia (variaţia inversă) a datelor senzoriale este mare; în consecinţă, predicţia posterioară trece decisiv spre dreapta. Dacă precizia datelor senzoriale ar fi mai mică, predicţia posterioară s-ar deplasa mai puţin sau chiar deloc.

Obiectivul nostru în interacţiunile cu lumea trebuie să fie *precizia.* De aceea, modelele noastre trebuie să aibă un mecanism pentru *predicţia preciziei.* Asta ne permite „să cântărim” precizia preconizată a semnalelor de eroare primite în ce priveşte precizia pe care am atribuit-o unei predicţii de ieşire. Acest lucru (gradele relative de încredere) va dicta influenţa semnalelor de eroare reale asupra predicţiilor noastre. Dacă avem tot mai multă încredere într-un semnal de eroare primit, ar trebui să avem tot mai puţină încredere în planul nostru curent de acţiune, iar evoluţiile nepreconizate vagi ar trebui să nu ne îndepărteze de la cursul nostru predeterminat. (Se pot atribui valori de încredere relativă şi aşteptărilor active versus aşteptările perceptive şi celor exteroceptive versus cele interoceptive, precum şi tuturor celorlalte cantităţi implicate în legea lui Friston[4].) Aceasta este un fel de inferenţă bayesiană de ordinul doi, care atrage după sine *inferenţe despre inferenţe,* adică niveluri de încredere informată despre predicţii.

Scopul modulării preciziei este de a garanta că inferenţele făcute de modelele predictive sunt influenţate de semnale de învăţare *fiabile* (ştiri demne de încredere): dacă există o încredere mare într-un semnal, atunci trebuie să i se permită să revizuiască o ipoteză anterioară şi invers dacă încrederea e mică. Încrederea afectează forţa semnalelor de eroare ce sunt propagate spre interior prin ierarhie. Un semnal în care ai încredere mare (adică unul mai precis) este mai „sonor”. Prin urmare, va avea o şansă mai mare să transmită o eroare reziduală spre centrul sistemului şi

să-şi actualizeze modelul generativ. În schimb, semnalele mai puţin precise – în care ai mai puţină încredere, cunoscute şi ca „zgomot” – pot fi izolate în epiteliul senzorial şi ignorate fără riscuri (aşa sperăm).

Asta înseamnă că trebuie să minimizăm semnalele de eroare precise. Din nou, pare paradoxal, până ce realizezi că nu e vorba decât de ideea că trebuie să evităm să facem greşeli monumentale. Asta nu se poate realiza decât prin îmbunătăţirea modelelor noastre generative, crescând astfel informarea reciprocă între modelele noastre despre lume şi eşantioanele senzoriale obţinute din ea. Cu alte cuvinte, trebuie să maximizăm precizia predicţiilor noastre, apoi să căutăm date precise care le confirmă. *Trebuie să ne maximizăm încrederea în credinţele care ne ghidează acţiunile,* ceea ce se numeşte „optimizarea preciziei”.

Realizăm optimizarea preciziei învăţând din experienţă. Trebuie să învăţăm în ce surse de informare putem avea încredere (şi când) şi să ne ajustăm predicţiile în consecinţă. De exemplu, avem încredere în semnalele vizuale ziua şi în semnalele auditive noaptea. Acordăm o pondere mai mare celor văzute ziua şi celor auzite noaptea, deoarece *am învăţat* să procedăm astfel. În consecinţă, ne aşteptăm ca informaţiile vizuale să fie mai precise ziua decât noaptea. Acest lucru se numeşte „precizie preconizată”. Pentru că nu ne aşteptăm la precizie vizuală noaptea, tolerăm fără ezitare tot felul de imagini neclare şi vagi, dar nişte experienţe vizuale similare în timpul zilei ne-ar face să credem că ceva nu e deloc în regulă.

La fel, angajaţii lui Eve Periaqueduct au învăţat că

presiunea în scădere în barometre prezice creşterea precipitaţiilor iarna, dar nu şi vara. Prin urmare, ei puteau să considere „zgomot” o valoare în scădere cu mai multă încredere vara decât iarna. Se *aşteptau* ca valorile barometrelor să fie mai puţin precise vara. În caz contrar, îşi ajustau nivelurile de precizie preconizate, care, apoi, le afectau predicţiile posterioare. Să dau un exemplu mai afectiv: oile din Sussex au învăţat să aibă încredere în predicţia lor că oamenii care aleargă spre ele ziua nu le fac rău, dar au mai puţină încredere în aceeaşi predicţie noaptea. În consecinţă, îi atribuie mai puţină precizie preconizată în timpul nopţii. Poate că, dacă vor continua să dea de mine în condiţii de întuneric şi se va dovedi că nu le fac niciun rău, îşi vor ajusta această valoare a preciziei, modificându-şi în consecinţă predicţiile despre oamenii care aleargă spre ele noaptea. Chestiunile dependente de context pot fi învăţate, ca orice altceva.

E clar din afirmaţiile mele că acest tip de învăţare se învârte în jurul *contextelor fluctuante.* În lipsa modelării predictive a dinamicii contextuale, un sistem cu autoorganizare nu poate supravieţui mult timp în mediile schimbătoare. Modelul generativ trebuie să încorporeze aceste dinamici. Trebuie să înveţe să anticipeze grade de precizie. Iar ajustarea valorilor preciziei, ca orice altceva în creierul predictiv, trebuie să respecte legea lui Friston.

Precizia este modul în care creierul îşi reprezintă gradul de încredere într-o anumită sursă de dovezi senzoriale sau în consecinţele preconizate ale unei anumite acţiuni. Valorile preciziei cuantifică aşteptările

în ceea ce priveşte variabilitatea. Aşadar, ele sunt *reprezentări ale incertitudinii.* Câtă încredere am în semnalul de eroare *x* în contextul curent? Câtă *pondere* să-i dau, chiar acum? 8/10 pentru *A* valorează mai mult sau mai puţin decât 8/10 pentru *B în condiţiile din acest moment?*

Deja am văzut fiziologia acestui lucru: triunghiul de decizie din mezencefal prioritizează o nevoie, apoi modelul sinelui în lume din prozencefal generează un context preconizat în care nevoia prioritizată va fi satisfăcută. Lumea preconizată are două faţete: conţinutul efectiv al predicţiilor şi nivelul de încredere a sistemului în predicţiile respective. Acum, că ştim cum sunt cuantificate nivelurile de încredere, le putem insera în descrierea pe care am făcut-o fiziologiei activării.

După ce triunghiul de decizie şi-a selectat nevoia predominantă curentă, ceea ce determină starea afectivă a sistemului, care determină apoi contextul preconizat generat de sistemele de memorie pe termen lung din prozencefal, sistemul reticulat activator se pune pe treabă. Sistemele mnezice atribuie *valori de referinţă ale preciziei* pentru contextul preconizat şi le aplică în toată ierarhia predictivă. Apoi un nor de neuromodulatori se vălătuceşte în prozencefal, impulsionând unele canale să se descarce rapid şi descurajând altele. Aceste frecvenţe de descărcare determină câtă pondere
se va atribui predicţiilor curente şi erorilor lor aferente, ceea ce determină cât de „sonor” se vor transmite erorile. Cu alte cuvinte, valorile preciziei determină

nivelul de încredere al sistemului în legătură cu efectele care preconizează că vor decurge din cursul acţiunii aflate acum în desfăşurare, la diversele niveluri ale ierarhiei. Apoi, din nou, speră la ce e mai bine şi se pregăteşte pentru ce e mai rău.

Multe lucruri surprinzătoare pot fi prezise odată ce te obişnuieşti cu ele. Totuşi, dacă nu putem decât să sperăm la ce e mai bine şi să ne pregătim pentru ce e mai rău, asta înseamnă că unele lucruri *nu pot* fi prezise. Iar aceasta este a doua parte a poveştii. Asta înseamnă că sistemul trebuie să-şi ajusteze nivelurile de încredere pe moment – adică *să moduleze* activarea în contextul evenimentelor aflate în desfăşurare, pe măsură ce ele se întâmplă.

Am afirmat mai sus că ne putem imagina că se dezvoltă un set extrem de complex de algoritmi ai modelului (indiferent cât de complicaţi devin aceştia), care calculează cererile relative de supravieţuire în toate situaţiile predictibile, şi că acesta îşi prioritizează opţiunile de acţiune în baza acestor calcule, indiferent de „explozia combinatorie”. Dar cum alege organismul între *A* şi *B* când *incertitudinea în sine* devine elementul principal ce determină selectarea acţiunii? Asta se întâmplă în situaţiile noi, de exemplu, care nu sunt în niciun caz rare în natură.

Fiziologii numesc „modulare a activării” ceea ce specialiştii în ştiinţe computaţionale numesc „evaluarea preciziei”[5]. Sunt acelaşi lucru. După cum tocmai ai văzut, un semnal de precizie nu este altfel decât ceea ce am numit semnal „sonor” în capitolul 6 – unul *puternic.*

Asta presupune că modularea încrederii într-un semnal de eroare trebuie să urmeze *deviaţiile de la forţa lui preconizată.* Iar aceste deviaţii trebuie minimizate. Ca în cazul tuturor semnalelor de eroare homeostatice, e „bine” (pentru noi, sistemele biologice) când lucrurile se întâmplă după cum am prezis şi „rău” când predomină incertitudinea.

Pe măsură ce se desfăşoară secvenţa de acţiune, nivelurile de încredere iniţiale sunt *ajustate* ascendent şi descendent de sistemul reticulat activator. (Să ne amintim de angajaţii lui Eve Periaqueduct care citesc şi ajustează instrumentele.) Cu alte cuvinte, contextul senzorimotor în desfăşurare este „palpat” – şi evaluările încrederii sistemului ajustate – pe baza fluctuaţiilor în desfăşurare în incertitudinea preconizată. Modificările la nivelul activării urmăresc coeficientul de încredere estimată al erorilor de predicţie testate. În acest mod, valorile de precizie fluctuante estimează coeficientul de încredere schimbător al semnalelor aflate în desfăşurare ce conţin noutăţile. La rândul lor, aceste valori determină tot ceea ce mai face sistemul, conform legii lui Friston.

Toate acestea sugerează că *optimizarea preciziei este baza statistico-mecanică a prioritizării semnalului în general* – adică outputul critic a tot ce am văzut că se petrece în triunghiul de decizie din mezencefal şi în sistemul reticulat activator. Optimizarea preciziei este modul în care numeroasele semnale de eroare ce converg în PAG au fost prioritizate de la început, aducând în conştientizarea afectivă nevoia

predominantă, ducând la o serie de opţiuni în curs de derulare într-un context preconizat, ghidat de precizii preconizate. Acum, acestea trebuie să fie modulate ţinând cont de evenimentele senzoriale nepreconizate.

Poate că toate acestea sună mai degrabă abstract. Din contră, eu cred că sunt descrieri a ceea ce se întâmplă concret. Atât de mult din experienţa noastră constă pur şi simplu din mici impulsuri senzoriale, când observi lucruri care nu sunt chiar cum te-ai aştepta să fie, urmate de sondări cognitive după moduri în care să elimini diferenţa. Îţi aminteşti că trebuie să trimiţi un e-mail: abia când mâna ta nu dă de ecranul telefonului îţi dai seama că deja îl căutai – dar, dacă telefonul nu e chiar acolo, lângă tine, atunci unde l-ai lăsat? În bucătărie, unde erai acum cinci minute?

Sau să luăm un exemplu mai semnificativ din punct de vedere afectiv: o întâlnire cu un posibil partener romantic. Crezi că l-ai putea (ai putea-o) cuceri diseară. Îţi imaginezi succesiunea posibilă a evenimentelor şi îţi stabileşti un plan de acţiune. Apoi speri la ce e mai bun. Nu eşti sigur ce-o să iasă, dar, în baza experienţei tale anterioare cu această persoană, îţi evaluezi şansele la aproximativ 7/10.

Pe parcursul serii, aspectele pe care te axezi (ceea ce este predominant) sunt destul de diferite de cele pe care te-ai fi axat dacă ai fi luat cina cu fratele tău. Şi tonul emoţional este diferit. Fiecare mic semn care sugerează că persoana aflată de cealaltă parte a mesei răspunde pozitiv la propunerile tale îţi provoacă valuri de entuziasm: tot mai multă încredere că planul tău funcţionează. Pe neaşteptate, partenerul la care speri

cască şi se uită la ceas. Ce înseamnă asta? Câtă pondere să atribui acestei evoluţii? Îţi fuge pământul de sub picioare. Oare ai interpretat greşit toate semnele anterioare? Analizezi fiecare mişcare şi gest. Cel mai mic indiciu suplimentar că sentimentele tale nu sunt împărtăşite te va pregăti pentru ce e mai rău şi vei iniţia planul B: îţi salvezi mândria prefăcându-te la fel de indiferent. Apoi privirile vi se intersectează. Oare asta înseamnă ce crezi tu că înseamnă? Da, aşa e! Apoi mâna îţi este atinsă uşor. Inima începe să-ţi bată mai tare. Se pare că, până la urmă, poţi să rămâi la planul A.

O nevoie prioritizată (în acest caz, DORINŢA SEXUALĂ) este sursa de incertitudine predominantă în acest moment. Inferenţele despre cauzele ei devin conştiente ca afect, deoarece fluctuaţiile nivelului tău de încredere referitoare la posibilele acţiuni necesare pentru a satisface această nevoie trebuie să fie modulate de sentimente. Sentimentele îţi spun cât de bine sau cât de prost te descurci. Contextul în desfăşurare ce generează fluctuaţiile trebuie să devină conştient de asemenea, din acelaşi motiv. Iată de ce am definit conştientizarea exteroceptivă a acţiunii şi percepţiei drept *afect contextualizat.* Acum avem o înţelegere formală, mecanicistă a ceea ce înseamnă asta. Totul nu e decât incertitudine resimţită.

Este crucial să observi că afirmaţia „Contextul în desfăşurare ce generează fluctuaţiile trebuie să devină conştient” explică de ce experienţa are aspecte duale. Nu e doar o chestiune de „mă simt aşa”, ci de „mă simt aşa *în legătură cu asta”.* Şi „în legătură cu asta” trebuie resimţit folosind o monedă comună (incertitudinea

aplicată) – deoarece contextul este principala sursă a incertitudinii referitoare la energia liberă. Economia minimizării energiei libere necesită o monedă comună.

Aceste fapte relevă că conştiinţa nu este doar o perspectivă subiectivă asupra dinamicii „reale” a sistemelor cu autoorganizare, ci şi o funcţie având propriile puteri cauzale determinate. *Sentimentul* unei nevoi (spre deosebire de simpla existenţă a unei nevoi) contează enorm pentru ce va face în continuare subiectul nevoii respective. Afectele orientează efectiv ce face un animal de la un moment la altul în condiţii de incertitudine. Percepţiile exteroceptive există exclusiv pentru că sunt resimţite *în relaţie cu* acţiunile orientate afectiv pe care le contextualizează.

Aceasta este funcţia centrală a unor lucruri precum atenţia. Concentrarea atenţiei funcţionează ca selectarea afectului, dar este aplicată lumii exterioare. Nevoia noastră de a reduce incertitudinea ne guvernează privirea, de exemplu, astfel încât mişcările rapide ale ochiului scanează zonele unei scene în care există o probabilitate mai mare să fie găsite informaţii mai precise[6]. Mai simplu, semnalele relativ puternice atrag atenţia: li se atribuie o precizie mai mare.

Aşa funcţionează predominanţa. Când sunt testate, caracteristicile „predominante” ale lumii minimizează incertitudinea referitoare la ipoteza prioritizată în mod curent a sistemului: ele sunt cele care, atunci când lucrurile se desfăşoară după cum s-a preconizat, ne maximizează încrederea în ipoteză. Astfel, agenţii activi sunt determinaţi să eşantioneze lumea ca să (încerce să) îşi confirme propriile ipoteze[7].

Întrucât acestea sunt, în fond, ipoteze despre cum să ne satisfacem nevoile, asta înseamnă că fiecare specie este determinată să-şi selecteze propria lume perceptivă. Orientarea perceptivă a fiecărei specii este dictată de lucruri care contează pentru ea. În consecinţă, oamenii, rechinii şi liliecii trăiesc în lumi (subiective) diferite. Percepi obiecte şi evenimente numai atunci când le observi, iar pentru fiecare specie sunt predominante altele. Nu poţi să vezi decât ceea ce testezi[8].

Biologul chilian Francisco Varela a exprimat acest lucru foarte bine: „Specia îşi prezintă şi îşi specifică propriul domeniu de probleme”[9].

Asta presupune că precizia nu poate fi determinată în mod pasiv – nu putem doar să aşteptăm ca să vedem ce semnale sunt puternice fără să avem niciun fel de aşteptare. Precizia trebuie să fie inferată, apoi *atribuită* de modelul generativ. Atenţia – care nu poate fi separată de precizie – poate să fie, în consecinţă, atât „captată”, cât şi „direcţionată”[10]. De exemplu, reducem în mod activ la aproape zero precizia în ceea ce priveşte erorile când ne culcăm, dar un eveniment suficient de surprinzător tot ne va trezi. În alte situaţii, e posibil să o amplificăm în mod activ, ca atunci când ne concentrăm intens asupra unui text dificil, pentru că bănuim că el conţine ceva important.

În articolul (2018) care a propus teoria noastră despre conştiinţă, am schimbat împreună cu Friston unele dintre simbolurile convenţionale folosite în ecuaţiile ce calculează energia liberă (vezi notele de

final ale capitolului 7). Le-am înlocuit ca să precizăm că păşim pe urmele lui Sigmund Freud, care a încercat în 1895 să ofere „o psihologie ştiinţifică, adică” să prezinte „procesele psihice ca pe nişte stări determinate cantitativ ale unor particule materiale detectabile”, astfel făcând aceste procese clare şi lipsite de contradicţii[11].

Acestea erau primele rânduri ale „Proiectului de psihologie”. Aici, Freud a folosit simbolurile *0*, *0*, *w* şi *M* pentru a indica patru sisteme ipotetice de neuroni responsabile pentru percepţie, memorie, conştiinţă şi, respectiv, acţiune; şi a folosit simbolul *Q* pentru a denota stimulii externi. Eu şi Friston am folosit această notaţie ca să denotăm vectorii echivalenţi în cadrul unui sistem care se autovalidează[12]:

Qq = stări externe, aşa cum sunt modelate de stările interne ale sistemului

0 = stări senzoriale

M = stări active

0 = predicţii

w = precizii

În plus, noi am folosit:

e = erori de predicţie (bazate pe *0* şi pe predicţia sa *0*)

F = energie liberă Friston (bazată pe *e* şi pe precizia *w*)[13].

De observat că niciuna dintre aceste cantităţi nu măsoară direct stările externe, deoarece sunt *ascunse* de un sistem cu autoorganizare. Asta înseamnă, în tot

ce urmează, că avem o descriere autonomă, independentă a dinamicii psihice în ce priveşte propriile stări interne ale sistemului *(Qq, w)* şi stările zonei Markov (*0*, *M*). Fiind acum echipaţi cu aceşti termeni, putem să formalizăm dinamica unui sistem ce se autovalidează în relaţie cu optimizarea preciziei.

Voi începe cu două ecuaţii care definesc energia liberă Friston în termenii cantităţilor pe care le-am prezentat. Prima ecuaţie spune că „energia liberă este (aproximativ) logaritmul negativ al probabilităţii de a întâlni unele stări senzoriale cu autor activ”[14]. A doua spune că „energia liberă preconizată scade (aproximativ) proporţional cu precizia logaritmului negativ”[15]. Aminteşte-ţi că dinamica sistemului care se autovalidează există ca să minimizeze energia liberă.

După ce am stabilit aceste relaţii, devine evident că de fapt există *trei* moduri în care un sistem ce se autovalidează reduce eroarea de predicţie şi, prin aceasta, minimizează energia liberă – nu doar cele două, evidente, pe care le-am descris mai sus:

1) Poate *să acţioneze* (adică să schimbe *M)* ca să modifice senzaţiile (*0*) astfel încât ele să se potrivească cu predicţiile sistemului. Aceasta este acţiunea.
2) Poate *să-şi schimbe reprezentarea lumii (Qq)* ca să genereze o predicţie mai bună (*0*). Aceasta este percepţia.

Iar acum, în plus:

3) Poate *să ajusteze precizia* (*w*) ca să se potrivească în mod optim cu amplitudinea erorilor de predicţie

primite (*e*).

Aceasta afirm eu că este conştiinţa[16].

Acest proces de optimizare finală, optimizarea *încrederii* sistemului, aşa cum este descris în textul de mai sus, este cel pe care eu şi Friston îl asociem cu evaluarea energiei libere ce stă la baza experienţei resimţite. Ecuaţiile care formalizează aceste dinamici sunt prezentate în notele de final[17]. Întrucât a treia ecuaţie este cea esenţială, o voi exprima în cuvinte: „În timp, rata schimbării preciziei (*w*) depinde de cât de multă energie liberă *(F)* se schimbă când schimbi precizia. Asta înseamnă că precizia va părea că încearcă să minimizeze energia liberă[18]. Rata acestui proces de minimizare a energiei libere este diferenţa dintre variaţie (precizia inversă) şi suma erorilor de predicţie la pătrat (*e.e*)”—. În termeni mai simpli, a treia ecuaţie cuantifică modul în care *ajustarea continuă a preciziei implementează legea lui Friston,* alături de acţiune şi percepţie. Pe scurt, cuantifică modul în care *conştiinţa contribuie la acţiune, la percepţie şi la actualizarea modelului* – şi, prin aceasta, la minimizarea energiei libere.

Figura 17 este o reprezentare grafică a acestor dinamici.

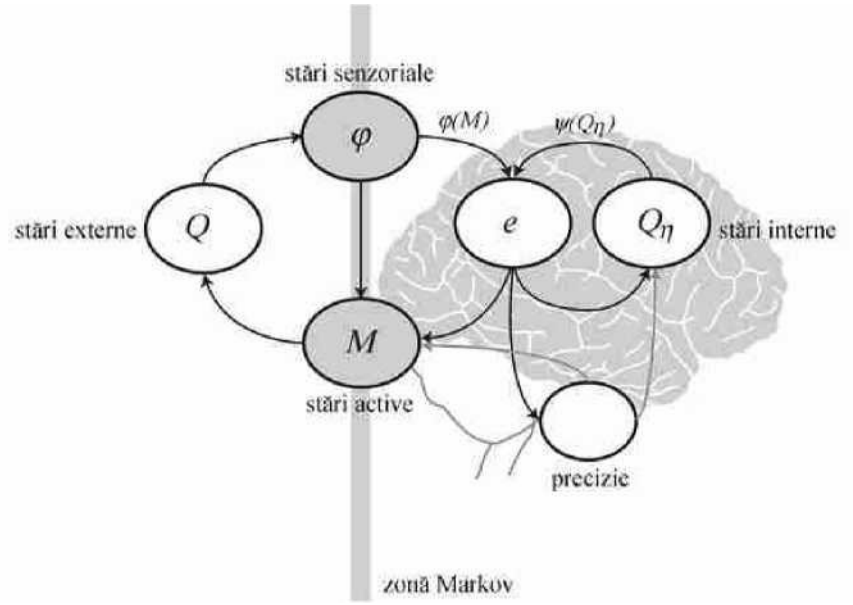


Figura 17. Dinamica unui sistem care se autovalidează, echipat cu optimizarea preciziei. Simbolurile sunt explicate în text. (*Q* este realitatea externă, care este ascunsă de sistem şi, prin urmare, nu apare în ecuaţii.)

La nivel conceptual, precizia este un determinant-cheie al minimizării energiei libere şi, prin urmare, al activării erorilor de predicţie. Precizia determină ce erori de predicţie sunt selectate – şi de aceea, în ultimă instanţă, modul în care reprezentăm lumea şi acţiunile noastre asupra ei. Dacă precizia este „activare” (ceea ce şi este), asta explică, din punct de vedere formal şi mecanicist, de ce optimizarea încrederii este întotdeauna şi exclusiv un proces endogen. Conştiinţa *nu are cum să vină decât* din interior[20].

Procesul pe care l-am prezentat aici se manifestă în multe moduri. Să-l mai privim o dată în termeni biologici (adică fiziologici şi psihologici). În domeniul *exteroceptiv,* se manifestă ca atentie şi atenuare, asociate cu creşterea şi scăderea preciziei senzoriale[21]. În domeniul *proprioceptiv,* corespunde preciziei potentialităţilor motoare (utilizări posibile ale obiectelor), de tipul asociat cu selectarea şi realizarea obiectivelor[22]. În domeniul *interoceptiv,* acest proces determină concret „senzaţiile viscerale”, deci cea mai bună explicaţie pentru semnalele interoceptive care au fost facilitate sau activate[23]. Totuşi, e foarte important să observăm că toate aceste lucruri (exterocepţie, propriocepţie şi interocepţie) pot să apară fără conştiinţă; conştiinţa este *simţirea* acestor lucruri.

Pe scurt: sarcina preciziei este *să activeze* reprezentări (şi aşteptări). În lipsa preciziei, erorile de predicţie nu vor reuşi să inducă nicio sinteză perceptivă sau un comportament motivat. Cu alte cuvinte, fără precizie, erorile de predicţie ar fi izolate acolo unde se formează. Asta s-a întâmplat, de exemplu, la pacienţii cu achinezie şi mutism ai lui Oliver Sacks.

Această formulare a preciziei implică mecanismele neuromodulatoare descrise în capitolul 6, care generează stări alterate de conştiinţă[24] şi vise[25] şi care sunt ţinta medicamentelor ce modifică conştiinţa (de exemplu, psihotropele şi psihedelicele)[26]. De asemenea, această formulare a preciziei dă o oarecare validitate versiunilor neuromodulatoare ale teoriei conştiinţei ca „spaţiu de lucru global” (discutată pe scurt în capitolul 4)[27].

Deloc surprinzător, rolul preciziei în psihopatologie este o temă importantă în domeniul în plină dezvoltare al psihiatriei computaţionale[28]. Să ne amintim de cazul domnului S., în care am văzut ce se întâmplă când semnalelor de eroare li se atribuie o pondere prea mică. Putem vedea din figura 17 cum anume triunghiul său de decizie şi, prin urmare, sistemul reticulat activator (w) au atribuit o pondere prea mare modelului său predictiv (Qq) şi prea mică erorilor sale de predicţie (e). După cum am văzut, de aici veneau sentimentele lui.

Ştiu că pare ciudat să vorbim despre conştiinţă în asemenea termeni mecanicişti. Dar am apelat la ei pentru că descriu legile aflate la baza experienţei fenomenologice, şi nu experienţa fenomenologică în sine. Prin prezentarea acestor legi, încerc să demonstrez că conştiinţa face parte din natură, că ea nu există în vreun univers paralel, că nu e ceva dincolo de ştiinţe.

Acum te rog să treci un Rubicon împreună cu mine.

Acest capitol abordează motivele pentru care şi modul cum apare conştiinţa, nu dintr-o perspectivă biologică, ci din una *formală şi mecanicistă.* Mai precis, pune două întrebări: (1) De ce şi cum se leagă procesul de prioritizare a nevoilor biologice pe care l-am descris mai sus de *legile* minimizării energiei libere? (2) De ce şi cum se întâmplă că acest proces mecanicist *face* unele sisteme cu autoorganizare să fie ceva cum e? Întrebarea de care m-am ocupat în ultimele câteva pagini, din perspectiva fizicii statistice, este cea dintâi. Aşa că acum trebuie să explic direct de ce şi cum

anume dinamicile statistico- mecanice pe care le-am descris generează experienţe resimţite. Cum poate un simplu proces de prelucrare a informaţiilor să simtă? Ca să explic acest lucru, trebuie să te rog să faci un salt pe care mulţi specialişti în ştiinţele naturale ezită să-l facă – iar din această cauză, ştiinţa în general şi mai ales ştiinţa psihicului au de pierdut. Mă refer la analiza dinamicilor mecanice pe care le-am descris *din punctul de vedere al sistemului.* Te rog să înlocuieşti perspectiva obiectivă, la persoana a treia din care am abordat până acum dinamicile în acest capitol cu una la persoana întâi: cu perspectiva *subiectivă* a sistemului care se autovalidează. Te rog să adopţi punctul de vedere al sistemului, *să empatizezi* cu el[29].

Acest salt este justificat de două fapte pe care le-am explicat deja. Primul este că experienţa resimţită poate fi consemnată numai din perspectiva subiectivă. A nu lua în considerare perspectiva subiectivă înseamnă aşadar a exclude din ştiinţă cea mai importantă trăsătură a minţii. Asta au făcut behavioriştii, pregătind terenul pentru jumătatea de secol în care neuroştiinţele nu au reuşit să abordeze conştiinţa într-un mod semnificativ. Al doilea fapt este: am arătat, în termeni formali şi mecanicişti, cum apare subiectivitatea sistemelor care se autovalidează. Prin urmare, asumarea perspectivei subiective a unui sistem care se autovalidează este justificată tocmai de faptul că posedă ipseitate.

Ca să fie clar ce spun: energia liberă şi preciziile sale constitutive pot fi cunoscute numai în interiorul unui sistem când el este înţeles în mod subiectiv, din punctul

de vedere al sistemului; experienţele nu pot fi observate *ca experienţe* din exterior, în mod obiectiv.

De asemenea, am explicat în termeni cauzali, principiali de ce şi cum ipseitatea unor astfel de sisteme este *intenţională.* Sistemele cu autoorganizare cu dinamicile pe care le-am descris au un obiectiv şi un scop – să supravieţuiască. Asta înseamnă că ele au un sistem de valori, cel aflat la baza tuturor formelor de viaţă.

Intenţionalitatea sistemelor dinamice cu autoorganizare le obligă să pună întrebări despre propriile stări în raport cu perturbaţiile entropice care le înconjoară. Asta le face sisteme „care se autovalidează". Ele *nu au cum să nu* întrebe mereu: „Ce se întâmplă cu energia mea liberă dacă fac asta?". Mai mult, sistemele complexe ce se autovalidează (ca noi, vertebratele) trebuie să pună această întrebare în relaţie cu *multe* variabile categoriale; deci răspunsurile – statisticile noastre vitale – trebuie să fie atât cuantificate, *cât şi calificate.* În sfârşit, ele trebuie să-şi moduleze nivelul de *încredere* în răspunsurile pe care le primesc.

Ceea ce descriu aici în termeni abstracţi, tehnici nu este prea complicat. O ştii din experienţă personală. Ceea ce trăieşti tot timpul sunt impulsuri fluctuante a ceea ce simţi ca răspuns la mişcarea ta prin lume, pe măsură ce verifici dacă totul e aşa cum te aşteptai să fie – şi pe măsură ce încerci să reduci cumva decalajul atunci când nu e cum te aşteptai. Practic, nu asta înseamnă experienţă pentru tine?

Prin combinarea tuturor acestor fapte despre

subiectivitatea şi intenţionalitatea sistemelor complexe care se autovalidează, ajungem la concluzia următoare. Răspunsurile evocate de echipament (în sensul lui Wheeler) care circulă în mod subiectiv pornind de la tipurile de întrebări la care sistemele ca noi sunt obligate să răspundă trebuie să aibă valoare existenţială şi mai multe calităţi. Încrederea noastră în aceste răspunsuri fluctuante – „fenomenele” – pe care le consemnăm trebuie să fie subiectivă, calificată şi să aibă valenţe.

Şi asta *înseamnă pur şi simplu* ceea ce înseamnă să experimentezi în mod conştient. Răspunsurile evocate de echipament, cel puţin în cazul vertebratelor ca noi (şi, fără îndoială, şi al altor organisme), sunt simţite.

Ca să te ajute să treci Rubiconul, aminteşte-ţi, te rog, că sentimentele *au evoluat.* Zorii conştiinţei au dat naştere unor *r* *r* *r* fenomene foarte simple, precum senzaţia că îţi este prea cald. Dovada senzorială tot mai precisă care prezice moartea unui sistem ce se autovalidează care se supraîncălzeşte pur şi simplu este resimţită de sistem ca „o căldură prea mare”. A trecut o veşnicie până s-au dezvoltat asemenea forme elementare de afect, în cadrul unei ierarhii predictive profunde, creând, în cele din urmă, „o lume tridimensională, panoramică, bine definită, compusă din obiecte solide cu formă clară: lumea experienţei noastre fenomenologice familiare”, ca în viziunea lui Merker.

Într-o asemenea lume, simţim ce înseamnă să fii un sistem cu dinamicile pe care le-am descris. Sentimentele sunt stări subiective fluctuante, cu valoare existenţială, cu calităţi diferenţiate şi cu grade de

încredere. Aceasta este materia conştiinţei. Acum înţelegem de ce nu are cum să fie altfel.

10
Înapoi la cortex

După cum am văzut în mod repetat în călătoria noastră, eroarea privitoare la cortex are de dat socoteală pentru multe. Dacă pionierii neuroştiinţelor behavioriste nu ar fi fost atât de impresionaţi de marea întindere a cortexului nostru sau dacă nu ar fi fost atât de orbiţi de ideea filosofică potrivit căreia viaţa psihică apare din asocierea unor imagini mnezice, poate că am fi descoperit de mult adevărata sursă a conştiinţei. În istoria ştiinţei psihicului, e o ironie ce nu ne dă pace că Freud avea atât de multe piese ale puzzle-ului acum mai bine de un secol. Avea sub nas indiciile, atât neurologice, cât şi psihologice. Dar, când a fost vorba de conştiinţă, până şi el a căzut pradă fixaţiei noastre colective pentru cortexul cerebral – obsesie al cărei preţ, în caz că uităm, poate fi măsurat nu doar în timp pierdut.

Toate acestea sunt adevărate. Totuşi, cortexul are în mod clar un rol masiv. Experienţa noastră de zi cu zi este strâns legată de dinamica procesării corticale. Aşa că, în acest capitol, să ne întoarcem la etajul superior înţeles greşit al creierului, ca să vedem ce adaugă el la

descrierea pe care o facem aici conştiinţei. Vom afla că multe dintre trăsăturile cele mai banale ale experienţei noastre de zi cu zi îşi derivă caracterul din ceea ce face cortexul – dar nu în modurile la care ne-am gândit până acum.

Sensul cel mai evident în care este valabilă afirmaţia de mai sus e că lumea aşa cum o resimţim se generează concret din reprezentări corticale. Deşi pare ciudat, în cadrul codificării predictive ceea ce percepem este o *realitate virtuală* clădită din propriile materiale de construcţie ale minţii.

Această afirmaţie este radicală în comparaţie cu perspectiva simţului comun, dar ideea că experienţa perceptivă este autogenerată se bucură de o largă recunoaştere în neuroştiinţele contemporane. De exemplu, să ne amintim ce a spus Semir Zeki despre vederea în culori încă din 1993. Zeki a scris că culoarea este *„o proprietate a creierului,* o proprietate cu care el învesteşte suprafeţele exterioare, o interpretare pe care o dă anumitor proprietăţi fizice ale obiectelor”[1]. Apoi a elaborat:

> Să presupunem că privim o mică suprafaţă scăldată izolat de o lumină cu unde lungi [...]. Suprafaţa produce un spectru larg de luminozitate pentru lumina cu orice lungime de undă, întrucât singura comparaţie pe care o poate face creierul în aceste condiţii este între lumina ce se reflectă din zona iluminată şi cadrul întunecat din jur. Astfel, lumina cu unde lungi produce luminozitate mare, iar cea cu

unde medii şi scurte, fiind absentă, nu produce luminozitate deloc. Astfel, sistemul nervos atribuie zonei culoarea roşie[2].

De observat cum îşi alege Zeki cuvintele: creierul *atribuie* lumii culoarea roşie, după ce a pus întrebări despre intensităţile relative şi lungimile de undă ale luminii. Face o pictură a lumii pe numere. Acelaşi lucru se aplică şi proprietăţilor fenomenologice care ne caracterizează celelalte modalităţi de percepţie: sunetele, gusturile, senzaţiile somatice şi mirosurile. Creierul atribuie lumii aceste calităţi.

Presupun că multor cititori le vine greu să creadă că ceea ce vedem chiar în acest moment nu este pur şi simplu ce e „acolo”. Îmi imaginez că întrebi: „De unde altundeva îmi vine perceperea cuvintelor de pe pagina asta?”. Poate te ajută dacă precizez că ce vezi acum nu prea seamănă cu inputurile senzoriale pe care le primeşti. La început, aceste inputuri sunt unde de lumină ce ajung pe retină. Aici, celulele fotosensibile (numite conuri şi bastonaşe) răspund la undele de lumină generând impulsuri nervoase. Aceste impulsuri – şi nu undele de lumină – sunt apoi propagate de-a lungul nervilor optici spre cortex sub forma unor *spike trains* (vezi figura 11). De ce resimţi aceste *spike trains* – 001111101101 – ca imagini mişcătoare aflate în lume?

Neuronii din corpul geniculat lateral şi din cortexul „de proiecţie” occipital ce răspund la impulsurile retiniene au o aranjare topologică, creându-se astfel posibilitatea formării unor imagini prin *cartografierea* suprafeţelor retiniene (vezi figura 6). Acest lucru însă

nu ţine cont de faptul că nu există bastonaşe şi conuri în apropiere de centrul retinei, de unde porneşte nervul optic. Atunci, s-ar justifica să vezi o gaură neagră mai spre centrul câmpului vizual. Cum dispare această gaură? Răspunsul e că *inferezi* din context şi din memorie ce este în „pata oarbă”, apoi o umpli[3].

Ar fi trebuit să spun că există o gaură neagră în câmpurile tale vizuale (la plural) pentru că, să nu uităm, ai două. De aici apare altă întrebare: de ce nu vezi două imagini? Nu mă refer la faptul că ai doi ochi. E simplu să ne imaginăm cum se pot suprapune două imagini aproape identice, dar nu asta se întâmplă[4], ci celulele din jumătatea stângă a fiecărei retine proiectează în lobul occipital drept, iar cele din jumătatea dreaptă proiectează în lobul stâng. Asta înseamnă că în cortexul vizual ai de fapt două reprezentări diferite ale suprafeţelor retiniene (una a jumătăţii din stânga a acestei cărţi şi una a jumătăţii din dreapta)[5], cu o fisură anatomică între ele: fisura longitudinală care împarte emisferele cerebrale. Cum anume devin cele două câmpuri imaginea unitară pe care o vezi? (E adevărat că sunt coordonate de axoni în corpul calos, dar şi oamenii cărora le lipseşte corpul calos văd o singură imagine[6].) În plus, trebuie să ţinem cont de faptul că câmpurile vizuale – aşa cum sunt ele reprezentate în lobii occipitali – sunt cu spatele în faţă şi cu susul în jos faţă de imaginile pe care le vezi. Mai mult, ochii tăi fug în toate direcţiile, cam de trei ori pe secundă, ca să nu mai vorbim de mişcarea constantă a capului. Cum se face că percepi o scenă vizuală stabilă şi orientată corect?

Ideea mea – că există puţine asemănări între ceea ce

vezi şi inputurile senzoriale ce ajung în cortex – este viu ilustrată de pacienţii cu antecedente neurologice cărora le sunt afectate mecanismele prin care transformăm în mod normal ceea ce ajunge acolo în ceea ce vedem. Eu am prezentat un asemenea caz cu mulţi ani în urmă: un băiat de 12 ani (W.B.) cu abcese bilaterale la lobii frontali, cauzate de pansinuzită[7]. Periodic, pacientul vedea lumea rotită la 180°. Simptomele şi semnele acestui pacient erau identice cu cele descrise în 21 de rapoarte ale unor cazuri anterioare pe care le-am găsit prin literatura de specialitate începând din 1805, conferind credibilitate descrierii

lui subiective (vezi figura 18 pentru unele dovezi obiective).

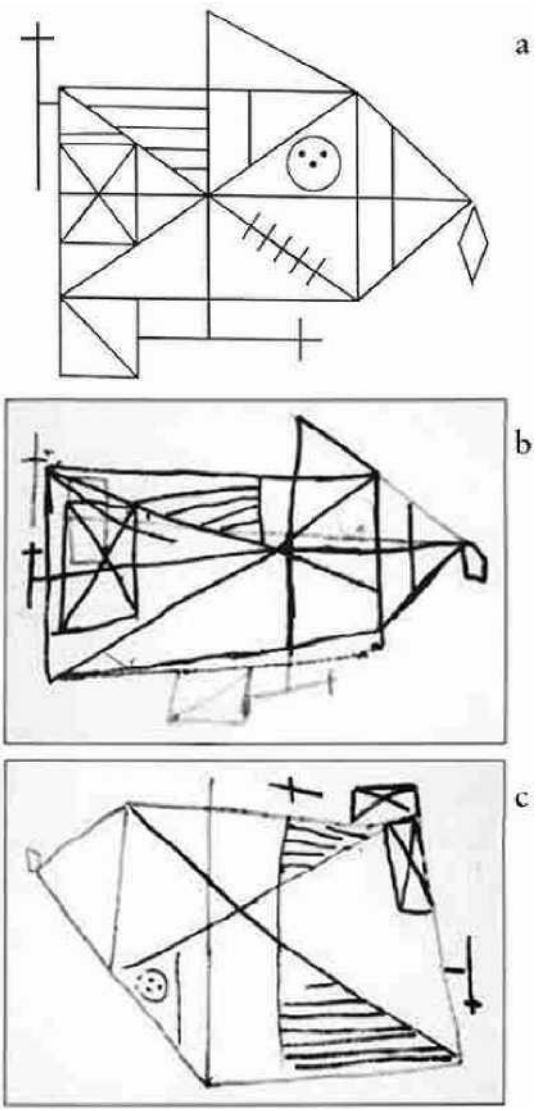


Figura 18. (a) = figura complexă Rey, aşa cum i-a fost arătată
pacientului W.B.; (b) = aceeaşi figură copiată de pacient; (c) =
figura desenată din memorie de pacient. Aceste desene constituie o dovadă obiectivă a inversiunii modelului predictiv
al lumii pe care îl avea W.B. Figura complexă din testul Rey este

dificil de reprodus chiar şi în cele mai bune momente;
de ce i-ar
creşte dificultatea acest pacient foarte bolnav,
desenând-o
răsturnată?

În prezent, mă ocup de o pacientă şi mai interesantă, care este inclusă într-un studiu de Aimee Dollman, o doctorandă a mea. Nu pot să prezint toate detaliile, deoarece cazul este încă în studiu, iar concluziile nu sunt publicate. Pacienta este o tânără foarte inteligentă cu disgenezie cerebrală (anatomie anormală) a lobilor occipitali, ce îşi reprezintă lumea aproape aşa cum tocmai am spus că *nu* o trăim: aşa cum este aranjată din punct de vedere anatomic în cortexul vizual. Asta se întâmplă mai ales când îşi foloseşte memoria vizuală (adică modelul ei predictiv). Vede două câmpuri separate, cu spatele în faţă şi cu susul în jos (nu întotdeauna la unison unul cu celălalt). Acest model vizual al lumii *nu reuşeşte să facă inferenţele corectoare obişnuite* prin care orientăm şi integrăm câmpurile vizuale. Prin urmare, acesta face predicţii eronate, iar experienţa ei vizuală nu se potriveşte cu cea a celorlalte modalităţi senzoriale. În consecinţă, uneori devine confuză în ce priveşte direcţia în care se mişcă corpul ei (mai ales când călătoreşte cu trenul sau cu avionul) şi face erori majore când se orientează prin mediul ambiant. În plus, îi vine greu să infereze care sunt obiectele constante şi care sunt datele vizuale pasagere (cum ar fi ortografierea unor cuvinte în minte când trebuie să le abstragă din fragmente şi fonturi

variabile sau identitatea unor chipuri abstrase în condiţii de lumină schimbătoare şi din perspective vizuale diferite).

Cu alte cuvinte, cortexul ei vizual primeşte tipul de informaţii pe care le primim cu toţii, dar ea nu poate să generalizeze pornind de la semnalele senzoriale sonore şi nici să infereze automat obiectele stabile reprezentate de acestea (de exemplu, să recunoască un chip familiar). Ea a suferit toată viaţa din cauza acestor anomalii şi şi-a dezvoltat metode elaborate prin care să le compenseze. Întrucât cortexul ei de „asociere” vizuală nu integrează *automat* cele două câmpuri vizuale, rotind scena, pacienta îşi ajustează reprezentarea lumii făcând inferenţe *deliberate.* De exemplu, când am rugat-o să identifice locaţia unui oraş bine-cunoscut pe o hartă nemarcată, a spus: „Să vă arăt unde *simt* că este sau unde *ştiu* că e?”. Atunci când „simte” că ceva se află la vest, ea „ştie” că se află probabil la est.

Totuşi, în loc să apelez la asemenea afecţiuni neuropsihologice rare, mai bine să ilustrez natura autogenerată a percepţiei cu ajutorul fenomenului utilizat de regulă în acest scop, şi anume *rivalitatea binoculară.*

Acest fenomen a fost descris pentru prima dată în 1593 şi a ocupat un loc important în opera fundamentală a lui Helmholtz pe tema inferenţei inconştiente[8]. Fenomenul implică prezentarea simultană a unor imagini diferite fiecărui ochi, cu ajutorul unui stereoscop cu oglinzi. Să spunem că ochiului stâng îi este prezentat un chip, iar celui drept o casă. În aceste

condiţii artificiale, experienţa vizuală se desfăşoară într-un mod „bistabil”, în care nu vezi o suprapunere a celor două imagini, ci o *alternare* între ele. Vezi o casă, apoi un chip, apoi o casă, apoi un chip, şi nu o imagine combinată casă-chip. Asta demonstrează clar deosebirea dintre semnalul obiectiv transmis creierului şi percepţia subiectivă generată de el. Helmholtz a concluzionat: „În asemenea cazuri, *interpretarea* [semnalului vizual] oscilează astfel încât observatorul are experienţe diferite, una după alta, pentru imaginea retiniană ce nu se schimbă”[9]. Prin urmare, ca şi în cazul vederii în culori, experimentezi o *inferenţă* despre inputul senzorial, şi nu inputul în sine.

În viaţa de zi cu zi e în general cam la fel, ca atunci când „am văzut-o” pe prietena mea britanică Teresa la aeroportul din Cape Town. Toate aceste iluzii arată că ceea ce percepi este generat în mare parte de *aşteptările* tale. În termeni bayesieni, se consideră că rivalitatea binoculară arată că, dacă *ipoteza anterioară* ce se potriveşte cel mai bine cu datele senzoriale (verosimilitatea mare că vezi o casă-chip) nu corespunde *cunoaşterii de fond* (probabilitate mică să existe case-chipuri), atunci ipoteza este respinsă. Inferenţa că vezi o casă domină inferenţa că vezi o casă-chip, aşa încât ceea ce vezi este o casă. Dar, atunci când testezi această ipoteză posterioară (ca o nouă ipoteză anterioară), ea se potriveşte numai cu jumătate dintre dovezile senzoriale. Cunoaşterea ta de fond îţi spune că chipurile întregi sunt la fel de probabile precum casele întregi. Aşa că te răzgândeşti şi inferezi că probabil vezi un chip – apoi asta vezi. Însă, atunci

când testezi această ipoteză nouă, ea nu explică, din nou, decât jumătate dintre dovezile senzoriale. Şi aşa mai departe...

Interpretarea bayesiană a rivalităţii binoculare este larg acceptată. Pe mine mă convinge pentru că, atunci când ni se arată două imagini care, în forma combinată, au o probabilitate anterioară mare, percepem o imagine amestecată. De exemplu, dacă ţi se prezintă două imagini astfel încât un ochi vede un canar, iar celălalt o colivie, le vei vedea imediat drept un canar într-o colivie.

Ceea ce percepi nu este totuna cu inputul care vine de la simţuri. *Ceea ce percepi este o inferenţă.* Iar materialele din care este derivată inferenţa sunt în cea mai mare parte modelul tău cortical predictiv derivat din experienţe trecute (adică preconizate)[10].

Asta ne spune ceva despre ce aduce cortexul în planul conştiinţei. Dar ce putem spune despre situaţia inversă? Ce face conştiinţa pentru cortex?

Ceea ce urmează să spun este evident, dar se pare că nu o mai exprimă nimeni[11]: conştiinţa cognitivă este generată de un mecanism neuronal recent descoperit numit „reconsolidarea memoriei”. La fel ca multe altele, această idee îşi are originea în scrierile lui Freud, conform căruia *„conştiinţa ia naştere acolo unde se sfârşeşte urma mnezică”*[12]. Ceea ce avea în vedere Freud era uşor diferit de ce urmează să spun eu, deoarece el – ca toţi neurologii vremii – era blocat în eroarea privitoare la cortex[13]. Totuşi, şi în acest caz e uimitor cât de aproape de adevăr a ajuns.

Am văzut că afectele înaintează cereri minţii, iar cogniţia face travaliul astfel solicitat. Mai exact, cogniţia *conştientă* face treaba; căci, odată ce sarcina a fost îndeplinită, iar încrederea în credinţa (prioritizată) ce devenise nesigură a fost restabilită, modelul generativ îşi reia modul automat de funcţionare, sub pragul conştiinţei[14]. Aici avem din nou mecanismul *învăţării din experienţă* pe care l-am tot descris. Acesta este rostul conştiinţei în cogniţie. Ajungi într-o situaţie în care nu eşti sigur ce să faci. Conştiinţa vine să te salveze: îţi cauţi ieşirea din scenariul respectiv în funcţie de ceea ce simţi, observând acţiunile voluntare ce dau rezultate bune în cazul tău. Apoi, treptat, lecţiile de succes devin automatizate, iar conştiinţa nu mai este necesară[15].

Vreau să evidenţiez că travaliul cognitiv pe care tocmai l-am descris *încetineşte* activarea în lume, care de altfel e ceva automat. Aceasta este diferenţa esenţială dintre acţiunea
voluntară şi cea involuntară, dintre cogniţia conştientă şi cea inconştientă, dintre pulsiunea resimţită şi reflexul autonom – ceea ce are caracter voluntar este mai puţin sigur şi, de aceea, necesită mai mult timp. Acest proces, care întârzie tendinţele de acţiune automatizată şi le permite să fie ţinute minte (în memoria pe termen scurt), se numeşte în mod inspirat „memorie de lucru”, care înseamnă, concret, că păstrezi în minte ceea ce simţi – afect stabilizat transformat în travaliu cognitiv. După cum tocmai am spus: dacă afectul înseamnă solicitare de travaliu înaintată minţii, atunci cogniţia

conştientă este travaliul în sine. Astfel, afectul însoţeşte şi totodată *devine* cogniţia. „Travaliul” despre care e vorba presupune inhibarea tendinţelor de acţiune automată şi *stabilizarea* intenţionalităţii în timp ce sistemul îşi caută ieşirea din problemele neprevăzute în funcţie de ceea ce simte. De aici vin avantaje adaptive considerabile, întrucât sunt facilitate soluţii viabile la numeroasele probleme din lumea reală pe care modelele noastre generative (încă) nu pot să le prezică. Acest proces stabilizator este funcţia cortexului[16]. Cortexul se specializează în incertitudine.

Ceea ce implică toate acestea este că *starea conştientă este indezirabilă* din punctul de vedere al unui sistem cu autoorganizare. Să revenim asupra diagramei de tip cadran din figura 12: săgeata din exterior reprezintă cererea tot mai mare de travaliu (afect negativ), iar cea din interior cererea în scădere (afect pozitiv), dar starea ideală este punctul de stabilizare ce reprezintă *zero cereri.* În capitolele 7-9, am oferit o fundamentare mecanicistă formală pentru aceste chestiuni. Minimizarea energiei libere este starea ideală a sistemelor vii, însemnând că surpriza minimă este idealul. Adică *nevoia* minimă este idealul. Afectul nu este decât anunţarea unei nevoi predominante. Asta ar trebui să însemne că sentimentul este ceva bun, deoarece permite sistemelor biologice să-şi rezolve nevoile şi, prin aceasta, să evite distrugerea. Dar starea *ideală* este cu siguranţă una în care toate nevoile noastre sunt satisfăcute automat – chiar înainte să fie simţite –, adică o stare *fără*

incertitudine. În această stare ideală teoretică, în care nevoile ne sunt satisfăcute automat, nu simţim nimic. (Aşa sunt satisfăcute majoritatea nevoilor noastre corporale: sunt reglate în mod autonom.) Spun „ideală teoretică” deoarece, în ce priveşte multe dintre nevoile noastre, în special cele emoţionale, nu o atingem niciodată. Numai pulsiunea CĂUTĂRII rezolvă acest lucru. Iar asta este vestea proastă.

Vestea bună este că semnalele de eroare cu cele mai mari valori de precizie au influenţa cea mai mare asupra modelului generativ. Întrucât cer schimbarea cea mai amplă – declară că faci ceva *greşit* –, ele constituie cele mai mari ocazii de a învăţa, astfel încât să nu fii la fel de încurcat data viitoare. E acelaşi lucru cu „a învăţa din greşeli”, după cum ţi-au spus să faci mereu părinţii şi profesorii. Acum ştii de ce: au vrut să te scutească de uitare. Dar efectul surprinzător rămâne că, în cogniţie, *conştiinţa este indezirabilă.* Prin urmare, toţi aspirăm nu la plăcere (nevoia în scădere), ci la starea de zombi (zero nevoi). Zero nevoi implică predicţii perfecte, adică zero erori şi, de aceea, zero cereri de creştere a preciziei pentru semnalele de intrare şi, prin urmare, nimic de simţit. În sfârşit, linişte.

Acum ar trebui să înţelegem mai bine aforismul lui Freud „Conştiinţa ia naştere acolo unde se sfârşeşte urma mnezică”. Cu alte cuvinte, *conştiinţa apare atunci când comportamentul automat duce la eroare* – când urma mnezică (o predicţie) care produce un comportament nu are efectul preconizat. Asta înseamnă că predicţia respectivă trebuie *actualizată* astfel încât

să înglobeze eroarea. Aşadar, conştiinţa corticală poate fi descrisă drept *„activitate predictivă în desfăşurare”.* O urmă mnezică ce este conştientă e în curs de actualizare. Nu mai este urmă mnezică. De aici apare conştiinţă în locul unei urme mnezice[17].

În ultimii ani ai secolului XX, am înţeles neurofiziologia acestui proces în termeni de „eroare de predicţie a recompensei” (vezi figura 19)—. Începând din primii ani ai secolului XXI, înţelegem mai bine fiziologia actualizării memoriei, numind-o „reconsolidare”.

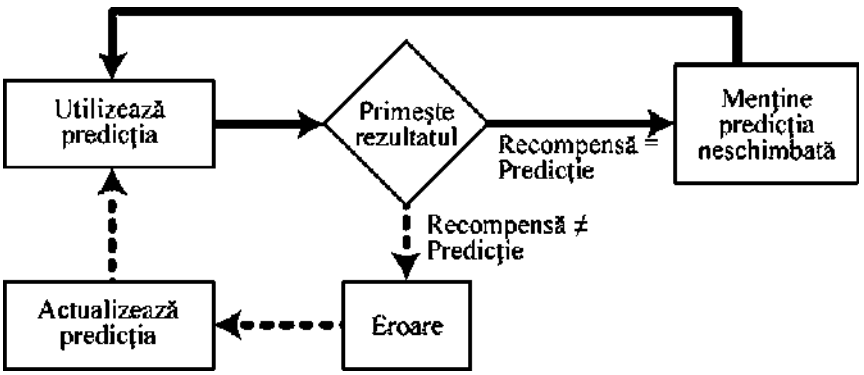


Figura 19. Schemă a învăţării prin eroarea de predicţie a recompensei. Secvenţa începe cu caseta „Utilizează predicţia”, care duce la două rezultate posibile, în partea dreaptă a casetei. Fie există o eroare de predicţie (când rezultatul diferă de predicţie), ceea ce duce la caseta „Actualizează predicţia”, fie nu există nicio eroare (când rezultatul se potriveşte cu predicţia),

ceea ce duce la caseta „Menţine predicţia neschimbată”, apoi comportamentul rămâne neschimbat. Etapa „Actualizează predicţia” corespunde reconsolidării.

Elementele fundamentale au ieşit pentru prima dată la lumină odată cu descoperirea, în anii 1960, a faptului că amintirea unei frici poate fi eliminată cu ajutorul terapiei electroconvulsive (ECT) – dar numai dacă electroşocul este administrat imediat după accesarea amintirii[19]. Asta sugera că ECT interferează cu un proces ce trimite în starea lor latentă amintirile manifeste (active) ale fricii: dacă electroşocul este administrat în timp ce amintirea fricii este activată, atunci această amintire este ştearsă. Concret, ea *nu mai este o urmă* în timp ce o rememorezi. Starea activată a amintirii face urma pe termen lung să fie *instabilă* din nou: o face să nu mai fie o amintire.

Acest lucru a fost confirmat în anul 2000, când reconsolidarea a fost redescoperită şi numită astfel[20]. Dacă inhibitorii sintezei proteinelor sunt administraţi în timp ce este activată o urmă pe termen lung, ea dispare. (Inhibitorii sintezei proteinelor împiedică formarea unor *noi* urme pe termen lung.) Asta e valabil şi în cazul altor tipuri de memorie, nu numai al amintirii fricii. În general, amintirile pe termen lung devin instabile când sunt în stare activată. În acest fel se actualizează (apoi se consolidează din nou – adică se reconsolidează).

În ultimele două decenii s-au realizat numeroase

studii privind reconsolidarea, arătându-se că ea apare nu numai la oameni şi la rozătoare, ci şi la găini, peşti, crabi, melci de apă dulce şi albine. Aceste studii arată şi existenţa a ceva analog reconsolidării în căile de procesare a durerii din măduva spinării, sugerând că are un rol primar în sistemul nervos central.

Memoria pe termen lung, spre deosebire de cea pe termen scurt, depinde de sinteza unor noi proteine, ce se declanşează după transmisia sinaptică substanţială şi repetitivă, care, la rândul ei, este modulată de sistemul reticulat activator – adică de activare[21]. De unde şi celebrul adagiu numit legea lui Hebb: „Neuronii care se descarcă împreună se şi conectează”[22]. O amintire „activată” nu mai este o amintire – se află într-o stare de incertitudine. Tot ce încerc să transmit este că conştiinţa cognitivă se reduce la inducerea instabilităţii la nivelul urmelor memoriei corticale şi că această instabilitate este un *produs al activării.* Tot ajungem din direcţii diferite la aceeaşi concluzie: procesele corticale sunt chestiuni fundamental inconştiente (sunt pur şi simplu nişte algoritmi, dacă sunt lăsate în voia lor). Conştiinţa – toată – vine din trunchiul cerebral[23].

Cu toate acestea, pare că ne străduim să atingem un fel de stare de zombi. Forma ideală de cogniţie este automatismul şi, cu cât reuşim să scăpăm mai repede de conştiinţă, cu atât mai bine. Şi atunci cum devine *inconştientă* cogniţia corticală?

Să ne gândim la cel mai simplu exemplu posibil. Ce crezi că s-ar întâmpla dacă aş proiecta o linie verticală în faţa ochilor tăi astfel încât să nu iau în calcul

mişcările constante ale ochilor, care, în mod normal, garantează că tot ce priveşti este actualizat cam de trei ori pe secundă? Această actualizare se petrece la periferia ierarhiei predictive, unde se găsesc cele mai mari grade de libertate. Ritmul frenetic în care se actualizează modelul necesar în asemenea circumstanţe impredictibile face cuvântul „predicţie” să-şi piardă sensul aproape în întregime. Iată de ce atât de mult travaliu cognitiv se desfăşoară la periferia senzorimotoare.

Aşadar, ce s-ar întâmpla dacă am putea să imobilizăm un stimul vizual în aşa fel încât să nu-l vedem decât pe el şi asta într-un mod cu totul monoton? Răspunsul tulburător este că *el dispare din conştiinţă.* Chiar dacă stimulul e tot acolo, el dispare din conştiinţa vizuală în câteva secunde. Acest lucru a fost demonstrat în anii 1950 de Lorrin Riggs în America şi de Robert William Ditchburn în Anglia. De atunci, au fost observate efecte similare şi la alte modalităţi senzoriale[24]. De-acum ar trebui să fie evident motivul pentru care stimulul dispare treptat din conştiinţă: el devine 100% predictibil, deci nu mai este purtător de informaţii. Predicţia atinge precizie totală, iar eroarea aferentă se reduce la zero. După cum am spus, acesta este idealul homeostatic al creierului.

E dificil să stabilizezi nişte stimuli în raport cu mişcările ochilor şi capului, aşa că nu poţi face acasă experimentul Riggs- Ditchburn. Dar poţi să cauţi pe Google iluzia „lilac chaser”, care prezintă alt tip de estompare vizuală[25]. După ce ai dat clic pe imagine, menţine-ţi privirea cam la 20 de centimetri de ecranul

calculatorului şi concentrează-te asupra ţintei din mijlocul cercurilor care se rotesc. Apoi observă ce se întâmplă cu cercurile. Ele dispar – deoarece creierul vizual inferează că sunt zgomot şi le atribuie mai puţină precizie în raport cu fondul gri, făcându-le să dispară treptat.

Pe site găseşti încă două iluzii optice. Pe lângă estompare, vei observa un cerc liliachiu care se transformă într-unul verde ce nu este acolo de fapt – (1) nici nu este cu adevărat prezent şi (2) nici nu e chiar verde – şi care se învârte în jurul ţintei centrale. Pe site se explică ce mecanisme funcţionează. Practic, e vorba de cântărirea preciziei. Toate aceste fenomene susţin ideea că ceea ce vezi e generat de creierul tău, nefiind primit din exterior[26].

Să identificăm legătura dintre cele de mai sus şi secţiunea din capitolul 5 despre actualizarea elementelor motoare anterioare înnăscute: reflexele şi instinctele. Aceste predicţii înnăscute ne ajută, dar ele nu au cum să descrie corect toate complexităţile lumii, aşa că trebuie completate prin învăţarea din experienţă. Învăţarea necesită conştiinţă, întrucât ne creşte treptat încrederea în predicţiile noastre nou-dobândite. Dar idealul oricărei învăţări este *să automatizăm* şi aceste predicţii dobândite, să le facem să se comporte precum reflexele şi instinctele. Sperăm să ne formăm predicţii noi care sunt cel puţin la fel de sigure şi de generalizabile ca şi cele vechi. În măsura în care putem realiza acest lucru, predicţiile dobândite se automatizează prin *consolidarea* sistemelor. În acest sens, consolidarea este opusul reconsolidării, care

anulează urmele consolidate, dizolvând la propriu proteinele care le-au „conectat”.

Asta se întâmplă până la nivelul sistemelor memoriei nondeclarative. Obiectivul învăţării din experienţă este trecerea cât mai multor amintiri pe termen lung din starea declarativă în cea nondeclarativă, pentru că „declarativ” înseamnă „capabil să revină în conştiinţă”. Aşadar, când am spus că starea de zombi este idealul cogniţiei, am vrut să spun că idealul învăţării este consolidarea tot mai profundă. Memoria nondeclarativă este *cea mai sigură* formă de memorie. Ea presupune cel mai mic volum de travaliu. Minimizează complexitatea şi este cea mai generalizabilă (vezi capitolul 8). În consecinţă, e cel mai rapid de executat; implică cel mai puţin incertitudinea şi, de aceea, întârzierea cea mai mică.

Desigur, există aici unele complexităţi. În primul rând, nu vreau să dau impresia că întreaga consolidare porneşte de la sistemele de memorie declarativă la cele de memorie nondeclarativă. Multe predicţii pe termen lung sunt consolidate *direct* în memoria nondeclarativă şi majoritatea tipurilor de învăţare au loc simultan în ambele sisteme. În al doilea rând, există multe tipuri de memorie nondeclarativă şi nu toate funcţionează la fel. De exemplu, învăţarea „procedurală” are loc prin repetiţie – prin urmare, spunem că abilităţile şi obiceiurile de tipul mersului pe bicicletă „se învaţă greu şi se uită greu”. Totuşi, unele tipuri de răspunsuri emoţionale nondeclarative, ce sunt la fel de greu de uitat, se dobândesc prin învăţare printr-o singură expunere: condiţionarea fricii, de exemplu. (Alte tipuri

de învăţare emoţională sunt mai lente; ataşamentul afectiv, de pildă, se face cam în şase luni.) „A fi greu de uitat” e trăsătura cardinală a memoriei nondeclarative, dar consolidarea presupune procese foarte diferite în numeroasele sisteme ale memoriei. În sfârşit, amintirile nondeclarative sunt „inconştiente” numai în sensul *cognitiv.* Atunci când este declanşat un răspuns emoţional dobândit, simţi ceva, doar că nu ştii despre ce este vorba – adică de unde a venit (vezi capitolul 3).

Cel mai important fapt despre memoria nondeclarativă e că este nondeclarativă. Ea generează *răspunsuri* procedurale, în timp ce memoria declarativă generează *imagini* experiate—. Acest lucru coincide cu o deosebire anatomică: amintirile declarative sunt corticale, iar cele nondeclarative sunt subcorticale—. Urmele mnezice subcorticale nu pot fi recuperate ca imagini pentru că nu constau din reprezentări corticale ale organelor senzorimotoare terminale. Ele presupun stereotipuri mai simple, de felul celor pe care le-am descris în legătură cu comportamentele învăţate ale copiilor cu hidroanencefalie şi ale animalelor supuse procedurii de decorticare cerebrală[29]. Asemenea lucruri nu pot fi aduse în conştiinţă, ele „nu pot fi gândite”.

În schimb, sistemele memoriei corticale sunt întotdeauna gata să ruleze scenariile predictive pe care le reprezintă – concret, să le retrăiască. Altfel spus, memoria declarativă livrează prompt urme pe termen lung stării pe termen scurt a memoriei de lucru conştiente. Face asta nu numai ca să actualizeze predicţiile corticale, ci şi ca să ghideze acţiunea în condiţii de incertitudine.

Urmele mnezice subcorticale sunt mai sigure decât cele corticale – în cazul valorilor lor mari de precizie, probabilitatea să se schimbe este mai mică – deoarece sunt optimizate pentru simplitate, şi nu pentru acurateţe. Asta le face mai generalizabile. Dar există un preţ: modelele mai puţin complexe sunt mai puţin precise când contextul variază[30]. Complexitatea relativă a predicţiilor corticale, în schimb, coincide cu mai multă plasticitate. Pe scurt, cortexul se specializează în *contexte* – reface acurateţea modelului în situaţii imprevizibile[31]. Un schimb este inevitabil. Cu cât e mai mare potenţialul de experienţă conştientă, cu atât mai redus e automatismul, ceea ce înseamnă mai multă plasticitate, dar şi mai mult travaliu cognitiv. Asta consumă energie şi generează sentimente, aşa încât creierul face minimul necesar. Chiar şi dacă asta înseamnă să estompeze până la anulare un stimul aflat sub ochii tăi.

Şi totuşi, o mare parte din ce se întâmplă în capul nostru pare greu de împăcat cu acest ideal de eficienţă informaţională şi termodinamică. Pe lângă sentimente şi percepţii, conştiinţa
noastră conţine, evident, *gânduri.* Ele au origine corticală, bineînţeles – dar ce sunt? Şi de ce par adesea atât de inerte?

Teoria cogniţiei pe care o prezint aici se învârte în jurul capacităţii sistemelor noastre mnezice de a genera o lume virtuală[32]. Fiecare nivel din ierarhia predictivă – inclusiv fiecare nivel al procesării corticale – este capabil să genereze o versiune a datelor pe care se

aşteaptă să le primească de la nivelul de după el. Asta înseamnă că percepţia nu e fundamental diferită de imaginaţie: din punct de vedere subiectiv, nu prea există diferenţe între lumile pe care le trăieşti în vise şi cea pe care o vezi la geam[33]. Creierul poate să evoce la comandă realităţi fantasmagorice. Probabil o face chiar acum, în timp ce citeşti aceste cuvinte şi te gândeşti la ce- or fi însemnând. Chiar în timp ce îţi laşi mintea să hoinărească...

Faptul că mintea poate să hoinărească ar părea o provocare pentru teoria conştiinţei din perspectiva energiei libere. Am spus că conştientizăm numai semnalele de eroare prioritizate, puternice – cele *predominante* –, la care trebuie să răspundem ca să ne putem păstra parametrii biologici în limitele viabile. Cu toate acestea, gândurile noastre par adesea să fie aleatorii şi nesemnificative. În unele cazuri, ni se pare chiar că monologul nostru interior ne deranjează sau ne distrage: nu ne ajută în situaţia în care ne aflăm. Cum poate aşa ceva să minimizeze energia liberă?

Deşi poate părea ciudat, scopul hoinărelii minţii este să ne îmbunătăţească *eficienţa* modelului generativ. După cum postulează principiul energiei libere, un model este eficient numai dacă foloseşte minimul de resurse necesare pentru a se autoorganiza. Asta se reduce la a găsi modelul cel mai simplu care prezice cu succes mostrele senzoriale din lume. (Să ne amintim de briciul lui Occam.)

Modelul cel mai simplu nu apare în mod natural dintr-o acţiune voluntară. Acţiunea voluntară este un proces aleatoriu. Prin urmare, simplitatea creşte prin

„toaletarea” conexiunilor sinaptice redundante ce s-au format când învăţam din experienţă. Le scurtăm ca să evităm „supraajustarea” modelelor noastre la date sonore, prin care am păstra inutil corelaţii excentrice şi slabe. Instrumentele cu care facem toaletarea sunt mecanismele de-acum familiare ale consolidării şi reconsolidării memoriei: prin activarea amintirilor, putem să le întărim, să le modificăm şi chiar să le ştergem.

Hoinăreala minţii este un mijloc prin care se realizează acest lucru. Implică o activitate spontană a prozencefalului (numită şi „stare de repaus” sau „mod implicit”), ce apare în lipsa oricărui stimul extern specific. Acest tip de activitate continuă în cea mai mare parte a timpului, în fundal, printr-o „explorare imaginativă a propriului spaţiu psihic”[34]. Între vis şi această formă de gândire există o suprapunere considerabilă[35], ce pare să survină la toate creaturile dotate cu cortex: orice animal cu capacitatea de a genera imagini ale sinelui acţionând în lume poate de asemenea să hoinărească prin nesfârşite lumi simulate, în funcţie de cum permit circumstanţele[36]. Poate îţi aminteşti că hoinăreala este strâns legată de pulsiunea CĂUTĂRII, ce îşi continuă cererile în timp ce dormim. Ar trebui să fie evident de ce activitatea în modul implicit este mai lipsită de riscuri noaptea, când nu e necesar să ne confruntăm cu evenimente exterioare.

Toate acestea explică faptul aparte că modelele lumii pe care le dobândim nu sunt niciodată stabile – nici chiar în timp ce dormim. Chiar şi în lipsa unor inputuri

senzoriale convingătoare, activitatea neuronală structurată continuă, generând cercetarea şi testarea neîntrerupte ale modelului generativ. Andy Clark tinde să creadă că aceste explorări ar putea duce la noi răspunsuri elegante la problemele care ne ocupă atenţia în starea de veghe – răspunsuri ce sunt adesea mai simple, în sensul lui Occam, adică *mai eficiente* decât cele mai bune dintre tentativele noastre anterioare: „Oare toate acestea ar putea fi cel puţin o parte din soluţia la enigmele profunde şi persistente referitoare la originile ideilor noi şi la rezolvarea unor probleme în mod creativ?”[37].

Conştiinţa joacă în acest proces autogenerat acelaşi rol ca în percepţie şi în învăţarea din experienţă. Toate procesele cognitive conştiente au în comun faptul că presupun travaliul psihic necesar al reconsolidării – întoarcerea predicţiilor consolidate la stările de incertitudine. Iată de ce visele (care sunt o formă de rezolvare a problemelor) sunt conştiente[38].

Dar există şi alte tipuri de gândire în afară de hoinăreala minţii. Să analizăm un al doilea tip: reveria *intenţionată.* Dacă există vreun proces cognitiv care merită rolul antagonist în opoziţia tradiţională dintre gândire şi faptă, atunci acesta este. Vorbim aici de gândire *în loc de* acţiune, care ne inhibă impulsurile motoare în timp ce sistemul caută în imaginaţie soluţii la probleme în funcţie de ceea ce simte. Când abordăm acest tip de gândire, ne ajustăm preciziile astfel încât erorile de predicţie motoare sunt suprimate[39]. Până la urmă, rostul gândirii deliberative este să-ţi imaginezi că

faci diverse lucruri ca să evaluezi *dinainte* care sunt consecinţele probabile ale realizării lor în realitate. (Am folosit în capitolul 5 exemplul directorului pe care îmi venea să-l lovesc.)

Cum ne imaginăm viitorul? Cam tot aşa cum ne amintim trecutul – care se dovedeşte a fi, mult mai frecvent decât ne place să recunoaştem, un trecut imaginat. Memoria episodică este un proces *construit* în care obiectivele şi contextele curente contribuie în mare măsură la ceea ce este readus în memorie. Astfel, trecutul este retrăit în mod selectiv şi tendenţios în funcţie de cererile curente de travaliu predictiv. Din nou, cazul domnului S. – căruia îi „lipsea un cartuş de memorie" – este un exemplu excelent, tocmai pentru că patologia lui îi exagera atât de mult mecanismele aflate la baza rememorării normale. Memoria lui episodică funcţiona alarmant de mult în propriul interes.

Bineînţeles, nu există cartuşe de memorie, dar sistemele neuronale implicate în călătoriile în timp la nivel mintal depind de *hipocamp,* care este crucial pentru introducerea aşa- numitului *„mine-ness"**2 perspectival în procesele mnezice corticale în mod normal inconştiente[40]. Studiile contemporane asupra memoriei episodice relevă faptul că hipocampul este la fel de implicat în imaginarea viitorului ca şi în retrăirea trecutului[41]. Din acest motiv, David Ingvar vorbeşte despre „amintirea viitorului"[42], iar Daniel Schacter consideră că hipocampul – împreună cu celelalte structuri cerebrale responsabile pentru memoria episodică – susţine o „simulare episodică constructivă"

a viitorului, care presupune „recombinarea flexibilă a unor detalii din evenimente trecute în scenarii noi”. În opinia lui Schacter, sistemul memoriei episodice îşi obţine valoarea adaptivă mai mult din capacitatea sa de a imagina viitorul decât din cea de rememorare a trecutului. Creierul, conchide el, este „un organ esenţialmente *prospectiv,* proiectat să folosească informaţii din trecut şi din prezent ca să genereze predicţii despre viitor”[43]. Toate acestea ar trebui să sune deja familiar.

Al treilea şi ultimul tip de gândire despre care voi discuta este gândirea cu *cuvinte.* Aparent, această capacitate este caracteristica remarcabilă a cogniţiei umane. Oamenii descriu limbajul drept un instrument de comunicare, ceea ce şi este. Dar înainte de toate e un instrument de abstractizare. Unii filosofi consideră că această „altă” funcţie a limbajului ţine de supracomunicare, dar eu prefer să consider că ţine de precomunicare – este dificil să ne imaginăm cum ar fi putut să apară vorbirea (spre deosebire de vocalizare) în lipsa abstractizării. Limbajul nu doar exprimă gânduri pe care le avem deja, ci formează gânduri noi.

Un experiment făcut de Gary Lupyan şi Emily Ward relevă această funcţie a limbajului. Cei doi au folosit o tehnică numită „suprimare continuă a instantaneelor” *(continuous flash suppression)* – alt tip de percepţie bistabilă, similară rivalităţii binoculare –, în care o imagine prezentată încontinuu unui ochi este suprimată din conştiinţă printr-un flux schimbător de imagini prezentate celuilalt ochi[44]. În acest experiment, subiecţii vedeau cu un ochi imaginea unui obiect familiar – cum

ar fi un scaun, un dovleac sau un cangur –, iar cu celălalt vedeau o serie de linii sinuoase. Liniile sinuoase au suprimat din conştiinţă imaginea stabilă. Totuşi, chiar înainte să privească liniile sinuoase şi obiectul, subiecţii au auzit unul dintre următoarele trei: (1) cuvântul denotând obiectul suprimat (de exemplu, „dovleac”, când obiectul era un dovleac), (2) cuvântul denotând un obiect diferit (de exemplu, „cangur”, când obiectul era un dovleac) sau (3) doar zgomot alb. Când li sa cerut să indice dacă au văzut ceva sau nu, posibilitatea ca subiecţii să raporteze că au văzut în mod conştient obiectul stabil atunci când cuvântul auzit se potrivea cu obiectul a fost semnificativ mai mare decât atunci când imaginea a fost însoţită de un cuvânt nepotrivit sau de zgomot alb. De fapt, auzirea cuvântului greşit a redus şi mai mult şansele de a vedea obiectul suprimat. Explicaţia este că, „atunci când informaţia asociată cu etichete verbale se potriveşte cu activitatea de intrare (de jos în sus), limbajul oferă un *impuls* de sus în jos percepţiei, propulsând în conştientizare o imagine altfel invizibilă”[45]. Cu alte cuvinte, ea creşte cântărirea preciziei pentru imaginea perceptivă.

Ca să înţelegem forţa acestui mecanism, să observăm că (în realitate, nu doar în acest experiment) ne furnizăm tot timpul asemenea etichetări de sus în jos, prin procesul „vorbirii interioare”. Impulsul astfel furnizat este o formă de *amorsaj* nondeclarativ. E interesant însă că amorsajul cu cuvinte are un efect substanţial mai puternic decât amorsajul cu imagini concrete. Se dovedeşte că o imagine valorează ceva mai

puţin decât un cuvânt, ca să nu mai zic de o mie. Asta se întâmplă probabil pentru că abstracţiunile (care se află mai în profunzimea ierarhiei predictive) pot să realizeze mai mult decât imaginile când e vorba de chestiuni precum „calitatea de a fi câine”. Astfel, de exemplu, Lupyan a descoperit că auzirea cuvântului „câine” are şanse considerabil mai mari să domine suprimarea continuă a instantaneelor decât simpla auzire a unui lătrat[46]. Abstractizarea are efecte mai profunde. Şi astfel, când subiecţii sunt amorsaţi implicit să fie atenţi la „vehicule” versus „oameni” în timp ce urmăresc un videoclip, amorsajul verbal modifică reglarea unor întregi populaţii neuronale, făcându-le să fie mai sensibile la prezenţa unei clase de obiecte versus cealaltă[47]. Cuvintele au puterea de a impulsiona categorii semantice întregi. Într-adevăr, multe asemenea categorii ar fi inimaginabile – şi, de aceea, imperceptibile – în lipsa etichetelor lor verbale. Tipul de concepte abstracte pe care le analizăm în cartea de faţă e cazul cel mai evident în care este valabilă această afirmaţie. A văzut vreodată cineva energia liberă? Totuşi, odată ce ne gândim la „energia liberă”, îi putem vedea mecanismele peste tot.

Dacă asta pot să facă unu-două cuvinte, ce se întâmplă când începem să alăturăm sute de cuvinte? De exemplu, să ne gândim la potenţialul naraţiunilor personale: povestirile abstractizate pe care ni le spunem despre curgerea şi sensul vieţii noastre. Andy Clark scrie că „asemenea naraţiuni funcţionează ca elemente de nivel superior în modelele care ne

structurează predicţiile despre sine şi, astfel, ne influenţează acţiunile şi aleg erile viitoare”[48].

Bineînţeles, aceste naraţiuni sunt *coconstruite* de obicei împreună cu alţii, pe parcursul vieţii, începând cu diada mamă – nou-născut. Aceasta introduce funcţia de comunicare a limbajului. Nu e neapărat necesar ca manipularea artificială a preciziei să vină numai din modelele noastre generative – ea poate veni la fel de bine şi din modelele altora, dacă aceştia au capacităţi similare de abstractizare. Clark include această a doua funcţie a limbajului în categoria „predicţie reciprocă continuă”, iar Andreas Roepstorff şi Chris Frith vorbesc de „partajarea scriptului” şi de „controlul de tip *top-top* al acţiunii”[49]. Pe scurt, etichetarea abstractă ce controlează precizia unei persoane poate fi comunicată direct preciziei alteia, evitându-se învăţarea laborioasă de jos în sus.

Roepstorff şi Frith argumentează această idee comparând efectele adresării de instrucţiuni verbale unor oameni care să-i ajute să îndeplinească o sarcină versus procesul anevoios al dresajului necesar pentru ca o maimuţă să înregistreze suficiente informaţii ca să facă acelaşi lucru. În exemplul folosit de Roepstorff şi Frith – un joc de sortare a unor cărţi de joc –, oamenii dobândesc comportamentul vizat după câteva minute de instrucţiuni verbale, iar pentru ca maimuţele să-l deprindă, e nevoie de un an întreg de condiţionare operantă. Ştim din dovezile furnizate prin RMNf că oamenii şi maimuţele din acest experiment au folosit regiuni cerebrale echivalente – acelaşi set cognitiv –

pentru *a îndeplini* sarcina. Diferită a fost doar metoda prin care *au dobândit* comportamentul.

Limbajul înseamnă chiar mai mult. El deschide calea multor tehnici diferite de îmbunătăţire a preciziei, cum sunt observaţiile, teoriile şi ecuaţiile prezentate în această carte. Datorită cuvintelor (şi altor simboluri, precum cele matematice), modelele dobândite pe parcursul unei vieţi pot deveni obiecte stabile care să fie analizate şi ameliorate sistematic de alţii – nu numai de contemporanii noştri, ci şi de generaţiile viitoare. Cu ajutorul minunatelor sale grade de generalitate şi de specificitate, limbajul ne permite chiar să proiectăm în conştiinţă ceva din structura ierarhiei predictive. Aceste ajutoare puternice ale cogniţiei nu sunt disponibile speciilor nonsimbolice. E foarte dificil să ne imaginăm ştiinţa, tehnologia şi cultura în general în lipsa limbajului.

Eu bănuiesc că limbajul a evoluat în principal din pulsiunea JOACĂ. Am explicat în capitolul 5 cum generează JOACA formarea regulilor sociale. Regulile reglementează comportamentul de grup şi, prin aceasta, ne protejează de nevoile potenţial excesive ale fiecărui individ. Regula 60:40 este o regulă socială înnăscută. Cere reciprocitate şi, de aceea, facilitează dezvoltarea empatiei – capacitatea de a citi alte minţi. E uşor de înţeles modul în care formarea regulilor sociale dobândite încurajează forme complexe de comunicare pentru a le exprima, precum şi modul în care acest lucru, la rândul său, contribuie la apariţia gândirii simbolice. Presiunea pentru a elabora reguli artificiale a crescut exponenţial când oamenii au trecut de la vechile

stiluri de viaţă de tip vânător-culegător caracteristice tuturor primatelor spre stilul de viaţă în aşezări permanente presupunând culturi agricole şi creşterea animalelor. Nu am fost pregătiţi sub aspectul evoluţiei speciei pentru această schimbare, ce a avut loc în urmă cu doar 12.000 de ani – singura asemenea „pregătire” o constituie pulsiunea JOACĂ, esenţială pentru formarea ierarhiilor sociale.

Prin urmare, este important să observăm că cortexul contribuie la JOACĂ mai mult decât la oricare dintre celelalte emoţii de bază[50]. Calitatea de „ca şi cum” a acestei pulsiuni este imposibil de conceput în lipsa mecanismelor corticale de tipul descris în acest capitol. JOACA ar putea fi chiar un precursor biologic al gândirii în general (adică al întregii acţiuni virtuale versus acţiunea reală) şi al întregii vieţi culturale.

Cititorii atenţi vor observa că afirmaţiile mele din această secţiune sugerează de asemenea ceva despre rolul limbajului în „terapia prin vorbire” în accepţiunea lui Freud. Ce este terapia prin vorbire dacă nu o intervenţie ţintită în naraţiunea cuiva? În opinia mea, psihoterapia – o formă de „predicţie reciprocă continuă” – este şi o formă de JOACĂ. Dar acum ating teme care merită dezvoltate aparte. Voi încheia aceste reflecţii despre cortex printr-o analiză a *diferenţei* esenţiale dintre conştiinţa corticală şi cea a trunchiului cerebral.

Panksepp era foarte interesat de tendinţa noastră de a asocia anumite culori cu anumite senzaţii[51]. De pildă, în mod obişnuit, descriem culorile din spectrul roşu drept calde, iar pe cele din spectrul albastru drept reci. Vorbim de ceva arbitrar? Oare culorile pot fi inversate?

Căldura şi răceala exteroceptive nu sunt, desigur, calităţi vizuale, cum sunt roşul şi albastrul, ci atribute ale senzaţiei somatice, fiind însă şi foarte apropiate de ceea ce Panksepp numea „afecte senzoriale”, precum dezgustul, durerea şi uimirea[52]. Din mai multe puncte de vedere, aspectele hedonice ale căldurii şi răcelii includ o valoare biologică – apropierea fizică de alţii este caldă, sexul este cald, focul este cald etc. E posibil ca spectrele de culori asociate cu focul şi gheaţa să-şi fi căpătat semnificaţiile afective în acest fel şi tot aşa, probabil, culorile fructelor coapte versus ale fructelor necoapte. Panksepp a speculat că aceste asocieri ar putea fi un vestigiu al originilor evolutive ale qualia perceptive, amintirea vremurilor când singurele noastre percepţii conştiente erau afectele senzoriale.

Dar există o mare diferenţă între simţirea afectelor şi atribuirea unor valenţe afective obiectelor percepute la un nivel exterior, cum ar fi fructele coapte sau focurile fierbinţi. Diferenţa este că afectele sunt înnăscute, dar atribuirea de valenţă obiectelor percepute externe este *dobândită.* Calităţile vizuale şi alte calităţi perceptive ale obiectelor percepute devin „asociate” cu afectele în sensul empirist[53]. Noi atribuim valori lumii. Deşi *unele* asemenea asocieri urmează regularităţi ce există în natură – între roşu, foc şi calitatea de a fi copt, pe de o parte, şi anumite senzaţii plăcute, pe de altă parte –, chiar şi aceste asocieri sunt dobândite şi, prin urmare, sunt supuse capriciilor experienţei individuale.

Pe scurt, asemenea asocieri sunt *contextuale.* De exemplu, focul nu este *întotdeauna* un lucru bun, motiv

pentru care, probabil, culorile din spectrul cald precum roşu uneori denotă lucruri bune, ca sexul, şi alteori lucruri rele, ca pericolul. Acesta este şi motivul pentru care vorbim nu numai de sex „fierbinte”, ci şi de conflict „fierbinte”. Acest principiu se aplică într-o măsură din ce în ce mai mare în cazul asocierilor tot mai individualizate care constituie iconografia subiectivă a fiecărei persoane. El funcţionează în logica Lisei Feldman Barrett.

Panksepp considera că valorile afective asociate cu unele calităţi perceptive sunt un vestigiu al originilor lor în afectele senzoriale. Dar nu trebuie să pierdem din vedere faptul că conexiunea înnăscută se rupe când facem legături între calităţi perceptive precum roşeaţa sau căldura şi trăirile erotice ori culoarea albastră sau răceala şi sentimentul de tristeţe. Iată motivul pentru care este atât de uşor să condiţionăm asocierile dintre culori şi afecte.

Regularităţile la nivel de populaţie ale acestor asocieri, care într-adevăr corespund naturii până la un anumit punct, trebuie să fie baza unor forme de artă precum pictura, muzica şi dansul. Le apreciem calităţile vizuale, auditive şi somatosenzoriale – cel puţin parţial – prin intermediul afectelor senzoriale şi al conotaţiilor pe care ni le evocă. Pentru că am definit conştiinţa drept incertitudine resimţită, e interesant de observat şi rolul surprinderii în experienţa estetică. Asta se aplică şi în cazul umorului: o operă de artă banală e ca un banc a cărui poantă o intuieşti dinainte. Dar aceste teme sunt prea vaste ca să le tratăm corespunzător aici[54].

Putem concluziona că aceste qualia perceptive sunt

diferite de cele afective în sensul că ele *nu au inerent* valenţe hedonice,
ci îşi dobândesc valenţa când sunt în relaţie cu afectul.

Lumina

Şi undele sonore ce bombardează mereu cortexul nu sunt resimţite întotdeauna conştient[55], dar, atunci când le conştientizăm, le simţim drept *context* (vezi pp. 151-152).

Până la urmă, percepţia constă din calcularea statisticii inferenţiale cu privire la distribuţiile de probabilitate ale aşa- numitelor *spike trains,* precum şi din compararea unor asemenea probabilităţi, toate într-o ierarhie în serie a homeostatelor ce optimizează precizia[56]. Ierarhia generează în ochii minţii noastre o reprezentare grafică a „răspunsurilor evocate de echipament" – *fenomene* în accepţiunea lui John Wheeler. Aceste fenomene constituie realitatea virtuală afişată înaintea fiecăruia (şi de fiecare) dintre noi. Stările de repaus ale miliardelor de mici homeostate – toate încorporate unele în altele, de la suprafaţă până la niveluri profunde – reprezintă *încrederea* noastră într-un context preconizat. Contextul este ceea ce ne aşteptăm să se întâmple dincolo de zona Markov, când încercăm să rezolvăm o nevoie prioritizată. Prioritizarea unei nevoi declanşează travaliu cognitiv înrădăcinat într-o credinţă aferentă: efectul preconizat. Ceea ce percepi conştient, prin mecanismul cu care suntem familiari de-acum, nu e acest context preconizat, ci afişarea în serie – într-o ierarhie profundă – a *devierilor* prioritizate de la aşteptările tale.

La periferia ierarhiei, asemenea fluctuaţii survin tot

timpul; e aproape imposibil să prezici prezentul în fiecare detaliu. Încrederea noastră fluctuantă în aspectele predominante ale aşteptărilor noastre senzorimotoare este resimţită sub forma culorilor, tonurilor şi altora asemenea – adică în aranjamente ale variabilelor categoriale exteroceptive calculate de specia noastră. Acestea includ nu numai perceperi ale lumii exterioare, ci şi devieri fluctuante de la stările proprioceptive preconizate ale avatarului ce îmi reprezintă corpul, întrucât felul cum îmi simt corpul nu e mai puţin virtual decât felul cum simt lumea exterioară[57].

Erorile de predicţie a preciziei în percepţie şi în propriocepţie (când sunt predominante) se consemnează drept qualia exteroceptive. În schimb, încrederea schimbătoare în credinţa fenotipică ce a generat de la început procesul senzorimotor este resimţită sub formă de afect. Indiferent ce ne propunem să realizăm, simţim în fiecare moment cât de bine ne descurcăm. Cu siguranţă că şi tu simţi asta acum, nu-i aşa? O fluctuaţie în interiorul tău pe măsură ce citeşti aceste cuvinte, un flux de incertitudine oscilantă, care te-ar putea opri dacă devine prea puternică. Calitatea sa fundamentală, afectul, este resimţită pe tot parcursul vieţii tale şi ea reglează până la urmă tot ceea ce faci.

Aş zice că experienţa noastră de zi cu zi nu constă până la urmă decât în asta.

*1 Sigmund Freud, „Dincolo de principiul plăcerii”, în *Opere esenţiale,* vol. III: *Psihologia inconştientului,* traducere de G. Purdea şi Vasile Dem. Zamfirescu.

Bucureşti: Trei, e-book (n.tr.).

*2 Faptul de a fi al meu (n.tr.).

11
Problema dificilă

Fizicianul Paul Davies scrie:

> Dintre numeroasele proprietăţi complicate ale vieţii, fenomenul conştiinţei iese în evidenţă în mod special. Originile sale sunt fără îndoială cea mai dificilă problemă cu care se confruntă ştiinţa astăzi şi singura ce rămâne aproape de nepătruns chiar şi după două milenii şi jumătate de deliberări [...]. Conştiinţa este problema numărul unu a ştiinţei, a existenţei chiar[1].

Aş putea cita multe afirmaţii similare făcute de alţi specialişti. „Problema dificilă” (cum e numită pe scurt şi respectuos) întreabă de ce şi cum anume a fost posibil ca tu – „bucuriile şi tristeţile tale, amintirile şi ambiţiile tale, sentimentul identităţii tale personale şi al liberului-arbitru”[2], pe scurt, experienţa existenţei tale – să izvorăşti din procesele fiziologice care au loc în celulele cerebrale. Aceste celule nu sunt fundamental diferite de cele din care sunt alcătuite alte organe. Deci cum fac

ele ca „tu” să existe?

Întrebarea nu e deloc nouă, ci e probabil cel mai vechi dintre toate misterele care îi preocupă pe oameni. Pe vremuri avea forma „Cum a ajuns sufletul meu să sălăşluiască în trupul meu?”, dar în forma sa curentă a fost exprimată în 1995 de filosoful David Chalmers. Iată celebra lui formulare:

> E incontestabil că unele organisme sunt subiecte ale experienţei. Dar întrebarea cum se face că aceste sisteme sunt subiecte ale experienţei ne pune în încurcătură. De ce, când sistemele noastre cognitive prelucrează informaţii vizuale şi auditive, avem experienţe vizuale sau auditive: calitatea albastrului-închis, senzaţia notei do din octava a patra? Cum putem să explicăm de ce există *ceva cum e* să ai o imagine mentală sau să simţi o emoţie? În general, se consideră că experienţa apare de la o bază fizică, dar nu avem nicio explicaţie satisfăcătoare a motivelor pentru care şi a felului cum apare astfel. De ce ar trebui ca procesarea fizică să dea naştere unei vieţi interioare bogate? Dintr-o perspectivă obiectivă, pare iraţional să existe această necesitate şi totuşi ea există[3].

Formularea lui Chalmers se inspiră în mare parte din ideile unui articol al filosofului Thomas Nagel, „What is it like to be a bat?” („Cum e să fii liliac?”, 1974). Nagel a subliniat acel *ceva- cum-e* al experienţei subiective. A arătat că „un organism are stări mintale conştiente dacă şi numai dacă există ceva cum e să *fii* acel organism –

ceva cum e *pentru* acel organism”, adăugând: „Dacă admitem că o teorie fizică a minţii trebuie să explice caracterul subiectiv al experienţei, atunci trebuie să acceptăm că nicio concepţie disponibilă în prezent nu ne oferă vreun indiciu despre cum s-ar putea face acest lucru”. Nagel a concluzionat: „Pare improbabil să putem lua în calcul vreo teorie fizică a minţii până nu analizăm mai mult problema generală a ce înseamnă subiectiv şi obiectiv”.

Două decenii mai târziu, Chalmers a reafirmat „problema generală” în termenii respectivi pentru că specialiştii în ştiinţele creierului tocmai începuseră să analizeze conştiinţa la un nivel experimental. Datorită progreselor tehnologice pe care le-am descris în capitolul 1, ei credeau că acum se poate răspunde la întrebări precum „În ce mod transformă celulele cerebrale în experienţe procesele fiziologice?”. Unul dintre primii cercetători care au abordat problema la nivel experimental a fost sir Francis Crick, specialist în biologie moleculară şi membru al echipei ce a descoperit structura ADN- ului. El a vorbit despre asta în volumul *The Astonishing Hypothesis: The Scientific Search for the Soul,* pe care l-a publicat cu doar un an înainte ca Chalmers să-şi declare dubiile. Iată ce a scris Crick:

> Ipoteza uimitoare este că tu, bucuriile şi tristeţile tale, amintirile şi ambiţiile tale, sentimentul identităţii tale personale şi al liberului-arbitru, nu eşti *de fapt* decât comportamentul unui mare ansamblu de celule nervoase, cu moleculele lor

aferente[4].

Chalmers nu a contestat categoric ipoteza că conştiinţa are o bază fizică. După cum tocmai am spus, el însuşi a afirmat că, „în general, se consideră că experienţa apare de la o bază fizică”. A spus doar că „nu avem *nicio explicaţie satisfăcătoare* a motivelor pentru care şi a felului cum apare astfel”. Prin aceasta a reluat aserţiunea lui Nigel că „nicio concepţie disponibilă în prezent nu ne oferă *vreun indiciu* despre cum s-ar putea face acest lucru”. Crick afirmase că se poate găsi o explicaţie satisfăcătoare: că tehnologia disponibilă în prezent ar putea să identifice cu uşurinţă ceea ce numea „corelativele neuronale ale conştiinţei”. Credea că, dacă izolăm părţile anatomice ale creierului ce sunt necesare pentru conştiinţă, precum şi funcţiile lor fiziologice specifice, am putea rezolva în mod ştiinţific problema minte-corp. Crick ne-a recomandat să începem căutarea concentrându-ne numai asupra unuia dintre corelativele neuronale ale conştiinţei: procesele cerebrale care disting *vederea* conştientă de cea inconştientă. Probabil că de aici am putea extrapola la restul conştiinţei. Părea destul de logic. Sigur există un corelativ neuronal al experienţei vizuale, nu-i aşa?

După cum am văzut, deşi numai vederea *corticală* este conştientă, cortexul poate să proceseze stimuli vizuali şi în mod inconştient, la fel şi coliculii superiori. Prin urmare, conform lui Crick, problema conştiinţei vizuale este o întrebare simplă referitoare la ce se petrece în cortexul vizual când prelucrează informaţii *conştient,* dar nu se petrece când o face *inconştient,*

ceva ce nu se întâmplă nici în coliculii superiori. Din motivele prezentate mai sus, eu cred că a pornit cu stângul, ca să vorbesc dintr-o perspectivă anatomică; ar fi trebuit să se axeze pe trunchiul cerebral, nu pe cortex, pe afect, nu pe vedere. Rezervele lui Chalmers au fost însă mai importante.

Perspectiva lui Crick – care a devenit abordarea principală în neuroştiinţele cognitive – eludează ceea ce Chalmers numeşte partea „dificilă” a problemei minte-corp. Izolarea corelativelor neuronale ale conştiinţei este partea „uşoară”. Nu face decât să identifice procesele cerebrale specifice care *se corelează cu* experienţa şi nu explică cum o *cauzează.* Asta este partea dificilă a problemei: cum şi de ce *produc* activităţile neurofiziologice experienţa conştiinţei?[5] Cu alte cuvinte, cum anume *devine* materia minte? Conform lui Chalmers, deşi noi, specialiştii în neuroştiinţe, putem să explicăm cum se prelucrează în creier informaţiile neuronale în timp ce avem experienţe vizuale, asta nu explică modul în care procesele cerebrale respective *se transformă în* experienţă. După cum s-a exprimat John Searle, alt mare specialist în filosofia minţii: „Cum anume depăşeşte creierul obstacolul dintre electrochimie şi sentimente?”[6]. Această întrebare ne derutează la fel de mult şi când este formulată invers: cum se transformă lucrurile imateriale precum gândurile şi sentimentele (de exemplu, te hotărăşti să-ţi faci un ceai) în acţiuni fizice ca prepararea unui ceai?[7]

Amploarea acestei lacune explicative[8], cum o numesc filosofii, este bine ilustrată de „argumentul cunoaşterii”,

care sună cam aşa[9]. Imaginează-ţi o specialistă în neuroştiinţe oarbă congenital pe nume Mary, care ştie tot ce se poate şti despre corelativele neuronale ale vederii. Deşi ea poate să explice toate realităţile fizice ale prelucrării informaţiilor vizuale – până la nivelul celular, inclusiv impactul undelor de lumină asupra conurilor şi bastonaşelor fotosensibile, precum şi modul în care aceste unde sunt transformate în impulsuri nervoase, apoi aceste impulsuri sunt propagate prin corpul geniculat lateral spre cortex, apoi acestea sunt prelucrate acolo mai departe de coloane de neuroni bine ordonaţi, organizaţi în număr mare într-o diversitate de module de prelucrare a informaţiilor răspândite peste tot în scoarţa cerebrală, ale cărei multiple fluxuri de procesare vizuală specializată sunt bine înţelese –, ea tot nu ar şti *cum e* să ai experienţa vederii. Fiind oarbă din naştere, ea nu ar şti nimic despre calităţile resimţite a ceea ce înseamnă roşu şi ce înseamnă albastru – de exemplu care sunt, până la urmă, materia propriu-zisă a vederii conştiente. Asta nu numai din cauză că ea nu a avut niciodată experienţa unor asemenea calităţi, ci şi din cauză că nimic din *cunoştinţele* ei de anatomie şi de fiziologie despre corelativele neuronale ale vederii nu explică cum e să vezi. Dacă ar căpăta brusc capacitatea de a vedea, Mary ar învăţa ceva cu totul nou despre vedere – ceva pentru care nu a pregătit-o nimic din înţelegerea ei mecanicistă. *Prin urmare, realităţile fizice nu explică de ce şi cum există ceva cum e să vezi.* Ele explică numai de ce şi cum decodifică creierul informaţii vizuale: cum

vede *el*, şi nu cum vezi *tu*. Această aşa-zisă ireductibilitate la baza fizică a ceea ce filosofii numesc „qualia” – acel ceva-cum-e al experienţei subiective – este problema dificilă. Conform lui Chalmers, acesta este „misterul central al conştiinţei”[10].

De-a lungul timpului, această ireductibilitate percepută i-a făcut pe marii gânditori, nu în ultimul rând pe fizicianul-filosof John Locke, să tragă concluzia că experienţele conştiente *nu fac parte din universul fizic*[11]. Întrucât e clar însă că qualia experienţei există, aceşti gânditori le-au relegat într-o dimensiune *nonfizică* a realităţii, pe care (mulţi dintre ei) au descris-o drept „epifenomenalistă”—. Locke a fost cel care a subliniat în „argumentaţia spectrelor inversate” că este posibil din punct de vedere logic ca o persoană să experimenteze calitatea albăstrimii la fel cum eu experimentez calitatea roşeţii şi să o numească tot „roşu”, ca mine, deşi ceea ce vede în mod conştient e ceea ce eu aş numi „albastru”. Consecinţa acestei afirmaţii e că aşa ceva *nu schimbă lucrurile* – aceste două extreme ale spectrului de culori ar putea fi inversate imediat în relaţie cu qualia lor resimţite. Aşadar, conform lui Locke, conştiinţa nu este explicată de corelativele ei fizice – relaţia dintre mecanismul fizic (în acest caz, lungimile de undă relative ale luminii) şi qualia psihologice nu este una cauzală.

Astfel, conştiinţa devine ceva ce doar rezultă pe lângă lanţul de evenimente fizice din creier – un fel de produs secundar –, fără impact asupra structurii cauzale. Locke scria în secolul al XVII-lea, dar această

perspectivă nu este arhaică, ci e larg răspândită, nu numai în rândul filosofilor. După cum au afirmat recent doi specialişti reputaţi în ştiinţe cognitive: „Conştientizarea personală este asemenea curcubeului ce însoţeşte procese fizice din atmosferă, dar nu exercită nicio influenţă asupra lor”[13].

Dacă experienţele conştiente nu joacă niciun rol în funcţionarea lumii fizice, de ce mai există? Ce anume adaugă conştientizarea la prelucrarea informaţiilor ce are loc oricum, în mod inconştient? Ce rost are să devii conştient de procesele cerebrale, dacă conştientizarea nu are nicio influenţă asupra lor?[14]

Această afirmaţie nu e valabilă numai în cazul vederii. Una dintre dovezile experimentale cele mai convingătoare ale opiniei conform căreia conştiinţa e un simplu epifenomen al proceselor cerebrale este o observaţie a lui Benjamin Libet, care spune că decizia ta subiectivă de a iniţia o mişcare este precedată (cu aproximativ 300 de milisecunde) de unde cerebrale măsurabile ce anunţă declanşarea mişcării pe care crezi că ai iniţiat-o tu[15]. Cu alte cuvinte, iniţierea fizică a mişcării (în creier) începe înainte să te hotărăşti în mod conştient să te mişti. De fapt, nu „tu” o iniţiezi. În general, se consideră că această descoperire demonstrează că alegerea conştientă – „liberul-arbitru” – este o iluzie. Dacă nu există liber-arbitru, atunci ce mai rămâne din conştiinţă? După cum am văzut, răspunsul lui Crick este că tu, bucuriile şi tristeţile tale etc., nu exişti „de fapt”. Dacă am avea o explicaţie fizică adecvată pentru asemenea qualia, spune Crick, le-am

putea *suprima.*

Mi s-a părut întotdeauna imposibil să accept atât argumentul conform căruia qualia conştiente *există într-un univers paralel,* cât şi argumentul că ele *nu există deloc.* Şi aşa ar trebui să faci şi tu, pentru că tu eşti conştiinţa ta. Este, evident, absurd să pretinzi că eşti ca un curcubeu care nu exercită nicio influenţă asupra corpului tău fizic. La fel de absurd e să pretinzi că „tu" nu eşti, de fapt, decât comportamentul unor neuroni, deci nu exişti cu adevărat. Aceste afirmaţii sunt contrazise de fiecare moment al experienţei tale.

Afirmaţia că „tu" nu exişti contrazice clar şi una dintre cele mai celebre concluzii din întreaga gândire filosofică. După o viaţă întreagă în care a meditat la *ce anume înseamnă certitudinea absolută,* René Descartes (în „filosofia îndoielii") a ajuns la celebra concluzie că *singurul* lucru în legătură cu care nu trebuie să avem nicio îndoială este faptul că existăm: „Cuget, deci exist". Cu alte cuvinte: ai experienţe, deci exişti.

Aceasta este problema dificilă a lui Chalmers: întrucât există în mod clar un sine ce are experienţe, cum îl integrăm în concepţia noastră fizică despre univers?

Eu nu sunt filosof. După cum ai văzut, interesul meu pentru neurologia conştiinţei s-a născut şi s-a dezvoltat independent de literatura filosofică şi mărturisesc că o bună parte din câtă filosofie am citit mai degrabă mă derutează. Cu toate acestea, în cartea de faţă încerc să ofer un răspuns ştiinţific firesc la „problema dificilă". Nu am pretenţia că am risipit orice mister din

metafizica conştiinţei. După cum vom vedea, rămân încă întrebări profunde. Dar cred că am demonstrat că procesele naturale, ce se desfăşoară în funcţie de diferitele lor necesităţi, pot (pe parcursul evoluţiei) să genereze ceva foarte asemănător cu lumile particulare ale experienţei pe care le avem.

Ce relevanţă are asta pentru problemă, aşa cum o formulează Chalmers? Cum rezolvă explicaţia pe care o dau lacuna explicativă dintre lumile interioare şi cele exterioare?

Să pornim de la o observaţie simplă: legile ce guvernează funcţiile psihice precum percepţia, memoria şi limbajul sunt *abstrase* atât din date subiective, cât şi din date obiective. Ca să dau un exemplu, legea lui Ribot, o abstractizare ştiinţifică despre memoria pe termen lung, explică faptul observabil că memoria resimţită *intern* a ceea ce s-a întâmplat acum zece ani este consolidată mai sigur decât ce s-a întâmplat acum zece minute. Acesta este motivul pentru care persoanele în vârstă tind să uite mai degrabă evenimente recente decât evenimente petrecute cu mult timp în urmă. Afirmaţia e valabilă şi în cazul urmelor mnezice fiziologice, observate *extern:* cele vechi de zece ani sunt mai bine consolidate decât cele vechi de zece minute (în exact aceeaşi măsură)[16]. Aşadar, gradientul temporal din legea lui Ribot nu este psihologic sau fiziologic, ci de ambele feluri. De asemenea, să ne gândim la legea lui Miller, o abstractizare ştiinţifică despre memoria pe termen scurt, ce explică limitarea resimţită a capacităţii de a ţine minte: în orice moment, nu poţi reţine decât şapte unităţi de informaţie

(plus/minus două). Şi durata amintirilor pe termen scurt poate fi măsurată: de obicei, ele durează 15-30 de secunde[17]. Exact aceeaşi limitare a capacităţii poate fi observată la nivel fiziologic: presupune depleţiunea neurotransmiţătorilor[18]. Prin urmare, legea lui Miller, la fel ca şi legea lui Ribot, este atât psihologică, cât şi fiziologică.

Legi ca acestea, care, în principiu, pot fi cuantificate şi deci exprimate matematic, explică manifestările duale ale unei funcţii abstrase numite „memorie”. Cu siguranţă acelaşi lucru e valabil şi în cazul „conştiinţei”. Dacă fenomenele conştiinţei sunt lucruri naturale (ce altceva ar putea fi?), atunci şi ele trebuie să fie reductibile la nişte legi.

Această concluzie nu este radicală sau idiosincrazică. Întreaga ştiinţă cognitivă, care domină de la sfârşitul secolului XX domeniul în care m-am specializat, se întemeiază pe asta. Să ne amintim că în capitolul 1 am scris:

> Cel mai interesant [la „prelucrarea cognitivă a informaţiilor”] e că poate fi implementată cu tipuri foarte diferite de echipamente fizice. Astfel, reuşim să vedem natura fizică a minţii într-o lumină nouă. Ni se sugerează că mintea (concepută ca prelucrare a informaţiilor) este o *funcţie,* şi nu o structură. Din această perspectivă, funcţiile de „software” ale minţii sunt implementate de structurile „hardware” ale creierului, dar aceleaşi funcţii pot fi implementate la fel de bine şi de alte substraturi, precum calculatoarele. Astfel, atât creierii, cât şi

calculatoarele îndeplinesc funcţii de *memorie* (codează şi stochează informaţii) şi de *percepţie* (clasifică tipare ale informaţiilor de input prin compararea lor cu informaţiile stocate), precum şi funcţii de *execuţie* (execută decizii referitoare la ce este de făcut ca reacţie la asemenea informaţii).

Totuşi, am continuat imediat:

Aceasta este puterea a ceea ce a ajuns să se numească abordarea „funcţionalistă”, dar este şi punctul ei slab. Dacă aceleaşi funcţii pot fi îndeplinite şi de calculatoare, despre care presupunem că nu sunt fiinţe conştiente, atunci se justifică să reducem mintea la simpla prelucrare a unor informaţii?

Şi aşa ajungem la esenţa problemei dificile. Conform lui Chalmers, *problema dificilă a conştiinţei nu poate fi rezolvată prin explicaţii „funcţionaliste”:*

Problemele simple sunt simple tocmai pentru că se referă la explicarea abilităţilor şi funcţiilor cognitive. Ca să explicăm o funcţie cognitivă, nu trebuie decât să specificăm un mecanism care o poate îndeplini. Metodele ştiinţelor cognitive se potrivesc pentru acest tip de explicaţie, deci sunt adecvate pentru problemele simple ale conştiinţei. În schimb, problema dificilă este dificilă tocmai pentru că nu se referă la îndeplinirea unor funcţii. Problema persistă chiar şi atunci când este explicată îndeplinirea tuturor funcţiilor relevante [...]. Problema dificilă

> este dificilă şi aproape unică pentru că trece *dincolo de* problemele legate de îndeplinirea unor funcţii. Ca să înţelegem acest lucru, să observăm că, şi după ce am explicat îndeplinirea tuturor funcţiilor cognitive şi comportamentale în vecinătatea experienţei [...], tot mai poate să rămână o întrebare fără răspuns: *de ce îndeplinirea acestor funcţii este însoţită de experienţă?* O explicaţie simplă a funcţiilor lasă deschisă această întrebare [...]. De ce toată această prelucrare a informaţiilor nu continuă „pe întuneric", fără să simţim nimic în interior?[19]

În capitolul 4, am rezumat tentativele eşuate de specificare a unui mecanism care să poată îndeplini funcţia conştiinţei. Răspunsul lui Chalmers la prima dintre ele ne lămureşte cu privire la atitudinea lui faţă de toate:

> Crick şi Koch sugerează că oscilaţiile [gama sincronizate] sunt *corelativele* neuronale ale experienţei. Această afirmaţie e discutabilă [...], însă, chiar dacă este acceptată, rămâne întrebarea legată de *explicaţie:* de ce oscilaţiile creează experienţă? Singura bază pentru o conexiune explicativă este rolul lor în fixare şi stocare, dar întrebarea de ce fixarea şi stocarea ar trebui să fie însoţite de experienţă nu e discutată niciodată. Dacă nu ştim de ce fixarea şi stocarea ar trebui să creeze experienţă, nu ne ajută să spunem o poveste despre oscilaţii. Şi invers, dacă *am şti* de ce fixarea şi stocarea creează experienţă, detaliile neurofiziologice nu ar fi decât

cireaşa de pe tort. Teoria lui Crick şi Koch prinde pentru că *presupune* o conexiune între fixare şi experienţă, iar astfel nu poate face nimic ca să explice legătura[20].

Iată critica lui Chalmers la a doua descriere funcţionalistă importantă a conştiinţei, teoria „spaţiului de lucru global” a lui Newman şi Baars (1993):

> Am putea presupune că, potrivit lui Baars, conţinutul experienţei este exact conţinutul spaţiului de lucru. Însă, chiar dacă este aşa, nimic din această teorie nu *explică* de ce sunt experiate informaţiile din spaţiul de lucru global. În cel mai bun caz, teoria poate să spună că informaţiile sunt experiate deoarece sunt *accesibile global.* Dar acum întrebarea apare într-o formă diferită: de ce accesibilitatea globală trebuie să ducă la experienţă conştientă? Ca de fiecare dată, această întrebare de legătură rămâne fără răspuns[21].

Chalmers concluzionează:

> Metodele explicative obişnuite ale ştiinţelor cognitive şi neuroştiinţelor nu sunt suficiente. Ele au fost elaborate tocmai ca să explice îndeplinirea funcţiilor cognitive şi îşi fac bine treaba. Dar, aşa cum se prezintă, ele sunt capabile *numai* să explice îndeplinirea funcţiilor. Când e vorba de problema dificilă, abordarea standard nu are nimic de spus […]. Ca să explicăm experienţa, avem nevoie de o abordare nouă[22].

Trebuie să clarific ce spune Chalmers aici. El afirmă că metodele ştiinţelor cognitive pot să explice numai îndeplinirea funcţiilor cognitive şi că îşi fac bine treaba. Prin „funcţii cognitive” se referă la lucruri precum percepţia, memoria şi limbajul. Aşadar, el spune că *conştiinţa nu este o funcţie cognitivă.* De ce spune asta? Răspunsul este: pentru că nu vorbeşte despre o *funcţie* abstrasă numită „conştiinţă”, ci despre *experienţa* conştiinţei – despre *cum e* să percepi sau să-ţi aminteşti.

Problema dificilă ar fi banală dacă s-ar reduce la faptul că experienţa ta individuală nu este identică cu experienţa umană în general[23]. Dacă asta ar fi problema dificilă, atunci, ca să o rezolvăm, nu ar trebui decât să facem o medie a experienţelor cu martor unic ale mai multor indivizi, să le găsim numitorul comun şi să explicăm *ce obţinem* în termeni funcţionali. Psihologii fac tot tipul asemenea lucruri. Asta am făcut şi eu când am studiat confabulaţia: am început cu experienţele subiective ale unui individ, domnul S., le-am generalizat prin studierea experienţelor echivalente ale altor pacienţi ca el, apoi am abstras numitorul comun. Această abordare a dezvăluit un principiu funcţional cu privire la confabulaţie, şi anume că îi face pe pacienţi să se simtă mai bine; confabulaţia are o funcţie „aspiraţională”.

Dar Chalmers nu spune doar că fenomenele subiective sunt conectate cu un singur punct de vedere. El întreabă: de ce ar trebui ca o funcţie, indiferent care, să fie însoţită de experienţă? Şi scrie:

Cum se face că, atunci când undele electromagnetice

> ajung la retină şi sunt discriminate şi clasificate de un sistem vizual, această [funcţie de] discriminare şi de clasificare este resimţită ca o senzaţie de roşu-aprins? Ştim că experienţa conştientă *chiar* apare când aceste funcţii sunt îndeplinite, dar însuşi faptul că apare este marele mister. Există o *lacună explicativă* (sintagmă pe care i-o datorăm lui Levine, 1983) între funcţii şi experienţă şi avem nevoie de o punte explicativă ca s-o depăşim. O descriere simplă a funcţiilor rămâne de o parte a lacunei, aşa că materialele pentru punte
> trebuie căutate în altă parte[24].

Merită să clarificăm ce trebuie să realizeze această „punte explicativă”, ca să nu pierdem din vedere ce ne interesează de fapt. Într-un articol publicat la scurt timp după ce eu şi Friston l-am predat pe al nostru (2018), trei colegi cehi au anunţat că e iminentă o soluţie la „problema dificilă”, prezicând că va veni de la principiul energiei libere:

> Nu există o perspectivă clară sau un consens general între filosofi şi specialiştii în neuroştiinţe asupra funcţiei conştiinţei. Acesta este unul dintre motivele principale pentru care conştiinţa încă e o problemă atât de greu de desluşit, ce îşi are rădăcinile în faptul că nu a fost formulată clar niciodată o funcţie a conştiinţei în baza unei teorii unitare a creierului, care să o şi susţină. O asemenea teorie unitară apare sub forma cadrului de codificare predictivă […] bazat pe ideile lui Hermann von Helmholtz, conform cărora

creierul este în principal un aparat pentru inferenţe predictive[25].

Totuşi, aceşti autori au adăugat un avertisment: „Din păcate, problema dificilă a conştiinţei probabil nu va dispărea niciodată cu totul, din cauză că va avea mereu susţinători devotaţi”. Şi au adăugat:

> Considerăm că sunt slabe şanse ca susţinătorii cei mai devotaţi ai problemei dificile să fie cu totul mulţumiţi de [orice] concluzii ale ştiinţei empirice, din moment ce, de la bun început, esenţa disputei legate de problema dificilă

vizează încercările ştiinţei empirice.

Bineînţeles, când eu şi Friston am primit primele evaluări din partea *Journal of Consciousness Studies,* exact asta ni s-a transmis despre articolul nostru: „Problema dificilă (conform lui Chalmers) este o problemă metafizică şi, ca atare, nu se pretează la a fi «rezolvată»”. (Friston mi-a scris ulterior: „Am impresia că problema dificilă nu există ca să fie rezolvată, ci ca să fie venerată”[26].)

Ca să fiu corect faţă de Chalmers, el nu este responsabil pentru prejudecăţile filosofice care îi preocupau pe colegii noştri cehi. Să ne gândim la ce spune el în fraza cu care îşi încheie articolul din 1995: „Problema dificilă este o problemă dificilă, dar nu există motive să credem că va rămâne veşnic nerezolvată”.

Aşadar, să spunem clar ce consideră el că trebuie să

realizeze puntea explicativă. Voi cita chiar criteriile lui Chalmers. Când a demontat prima teorie funcţionalistă, a scris: „Nu e abordată niciodată întrebarea de ce fixarea şi stocarea [sincronizate în frecvenţă gama] ar trebui să fie însoţite de experienţă”. Aşadar, când analizezi dacă lacuna explicativă este eliminată prin vreo teorie funcţionalistă a conştiinţei, trebuie să te întrebi: *este discutată întrebarea de ce funcţia XYZ ar trebui să fie însoţită de experienţă?* În ce priveşte a doua teorie pe care a respins-o, Chalmers a adăugat: „De ce accesibilitatea globală ar trebui să dea naştere experienţei conştiente? Ca întotdeauna, această întrebare de legătură nu primeşte răspuns”. Aşadar, când evaluezi orice altă teorie, trebuie să te întrebi: *se răspunde la întrebarea de ce funcţia XYZ ar trebui să genereze experienţă conştientă?* Şi se pare că asta-i tot.

Dar, dacă asta-i tot, atunci de ce adaugă Chalmers că „o simplă descriere a funcţiilor [experienţei] rămâne de o parte a lacunei, aşa că materialele pentru punte trebuie căutate în altă parte”? De ce trebuie ca o descriere a funcţiilor să nu reuşească să elimine lacuna? Dacă recitim ce spune el în acest sens, după cum l-am citat mai sus, devine clar că el combină *două tipuri diferite de lacună explicativă.* Aşadar, trebuie să spunem clar despre ce lacună vorbim, ca s-o putem elimina.

Prima este o lacună explicativă între semnalele propagate de la retină şi senzaţiile de roşu-aprins – adică între evenimentele fiziologice şi cele psihologice. Trebuie să evidenţiez că *ambele tipuri de evenimente sunt, de fapt, experimentabile.* Poţi trăi evenimente

fiziologice precum observarea semnalelor propagate de la retină la fel de simplu cum poţi trăi evenimente psihologice precum perceperea culorii roşu-aprins. Niciunul dintre aceste evenimente nu poate fi explicat – şi cu atât mai puţin suprimat – de celălalt. *Acelaşi lucru poate fi observat în două moduri.* Atunci când am experienţa subiectivă că exist, este cevaul psihic care există diferit de Mark Solms corporal pe care îl văd în oglindă? Şi este Mark Solms din oglindă ceva diferit de acel Mark Solms anatomic văzut printr-un aparat de RMN? Nu înţeleg de ce atât de mulţi oameni (chiar filosofi) vorbesc despre creier ca şi cum ar fi cumva exceptat de la realitate *aşa cum o experimentăm*[27]*.* Asta nu poate fi decât pentru că, atunci când spun „creierul”, ei se referă la altceva. Nu se referă la creierul pe care îl vezi şi îl atingi, ci la ceva *abstras* din experienţa pe care o avem despre el – ceva de natura unui sistem *funcţional.*

Aşadar, prima lacună explicativă este între *două tipuri diferite de experienţă* asociate cu două perspective observaţionale diferite[28]. Acest lucru e comparabil cu *a auzi* tunete cu urechile şi *a vedea* fulgere cu ochii. Lumea nu spune: „Se consideră în general că tunetul vine din fulger, dar nu avem o explicaţie satisfăcătoare a motivelor pentru care şi a felului cum se întâmplă aşa”. Asta e pentru că lumea nu crede că fulgerul *produce* tunetul aşa cum ficatul produce bilă. Se acceptă că ele sunt două manifestări ale aceluiaşi lucru[29]. Asta se aplică şi modurilor diferite de experiere a prelucrării informaţiilor vizuale: din exterior sau din interior. Din exterior (dacă eşti un

cercetător dotat cu echipamentul adecvat) vezi semnale propagate de la retină—; din interior vezi roşu-aprins.

A doua lacună explicativă este între experienţe (de ambele tipuri) şi cauzele lor fundamentale. E lacuna dintre ceea ce poţi experimenta, precum nuanţele de roşu-aprins şi imaginile unor neuroni activaţi obţinute prin metodele optogeneticii, şi ceea ce nu poţi experimenta, precum câmpurile cuantice. Pe scurt, e vorba de o lacună între perspectiva persoanei întâi şi perspectiva persoanei a treia. Să adopt o perspectivă la persoana a treia asupra propriei experienţe înseamnă să mă abstrag din experienţă, deci să nu o mai am. Această perspectivă nu vizează nici creierul aşa cum arată, nici mintea aşa cum se simte, ci forţele care *explică* de ce şi cum arată şi se simte aşa cum arată şi se simte. Aceasta este perspectiva pe care am adoptat-o în cartea de faţă.

Nu sunt sigur că este evident pentru toată lumea că Chalmers, în citatul de mai sus, se referă simultan la ambele lacune. Asta poate crea confuzie. El combină lacuna dintre două tipuri de experienţe (observaţii extrospective versus observaţii introspective) cu lacuna dintre experienţă în general şi mecanismele ei funcţionale fundamentale. Asta complică şi mai mult situaţia.

Mie mi se pare că majoritatea oamenilor presupun că ceea ce Chalmers numeşte „funcţie” este sinonim cu ceea ce numeşte „ceea ce e fizic” şi că „experienţă” este, prin urmare, întotdeauna şi exclusiv ceva nonfizic. Dar, dacă „ceea ce este fizic” înseamnă corpul observabil şi organele sale, inclusiv creierul, atunci nu este experiat mai puţin decât ceea ce este psihologic*[1].

Mai mult, experienţele psihologice, când sunt abstrase, dezvăluie mecanismul funcţional al unor asemenea experienţe, dând naştere unor legi psihologice.

Acelaşi lucru e valabil şi în cazul legilor fiziologice: şi ele sunt abstrase de la experienţă – de la datele fiziologice observabile. Aceste două tipuri de legi, fiind făcute din aceeaşi materie (explicativă) abstrasă – funcţiile –, nu sunt la fel de greu de redus una la cealaltă cum sunt cele două moduri de experienţă ce constituie categorii distincte. Astfel, legile lui Miller şi Ribot sunt atât psihologice, cât şi fiziologice şi, din acest motiv, pot fi reduse la nişte ecuaţii unificatoare. Faptul că nu sunt explicitate etapele intermediare dintre datele ce rezultă din experienţă şi mecanismele explicative complică problema dificilă şi o face să pară mai dificilă decât este. Nu e nevoie să-ţi imaginezi cum un tip de experienţă ar putea produce alt tip de experienţă: trebuie doar să găseşti o teorie mecanicistă care explică ambele seturi de fenomene, indiferent de modalitatea ei perspectivală. Apoi poţi să testezi predicţiile pe care le generează.

Chalmers admite că problema dificilă poate fi rezolvată. De fapt, el stabileşte trei principii pe care trebuie să se bazeze
soluţia. Înainte să prezint aceste principii (pe unul dintre ele trebuie să-l păstrăm pentru capitolul 12), trebuie să precizez că Chalmers consideră că nu vom putea rezolva niciodată problema dificilă dacă încercăm să o facem *reducând* „experienţa” la ceea ce el numeşte „procese fizice”. Chalmers scrie:

> Suntem deja în poziţia din care putem să înţelegem anumite realităţi esenţiale despre relaţia dintre procesele fizice şi experienţă, precum şi despre regularităţile care le conectează. *Odată ce lăsăm deoparte explicaţia reductivă,* putem să punem pe masă realităţile respective pentru ca ele să-şi joace rolul corespunzător, de piese iniţiale într-o teorie *nonreductivă* a conştiinţei şi de constrângeri asupra legilor de bază care constituie o teorie finală[31].

Pentru mine explicaţia „nonreductivă” este bună dacă înseamnă că putem renunţa la sarcina imposibilă de a reduce fenomenele psihologice la unele fiziologice sau invers. Fenomenele psihologice nu pot fi reduse la unele fiziologice, la fel cum fulgerul nu poate fi redus la tunet. Fulgerul nu *cauzează* tunetul; cele două fenomene *se corelează.* Aceasta este problema simplă. Prin urmare, trebuie să reducem ambele fenomene la mecanismele lor, astfel încât să putem reduce aceste mecanisme la un numitor comun, fără să încălcăm legile fizicii. Aceasta este problema dificilă.

Totuşi, Chalmers pare să se refere la ceva diferit prin termenul explicaţie „nonreductivă”. Pentru el, asta înseamnă că nu putem să reducem fenomenele psihologice trăite la nişte legi funcţionale şi atât.

Chalmers continuă:

> O teorie nonreductivă a conştiinţei va consta dintr-un număr de *principii psihofizice,* care conectează proprietăţile proceselor fizice la proprietăţile

experienţei. Putem considera că aceste principii includ modul cum apare experienţa din ce e fizic. În cele din urmă, aceste principii ar trebui să ne spună ce tip de sisteme fizice vor avea experienţe asociate, iar în cazul sistemelor care au, ar trebui să ne spună ce tip de proprietăţi fizice sunt relevante pentru apariţia experienţei, precum şi exact la ce tip de experienţă ar trebui să ne aşteptăm să genereze orice sistem fizic dat. Asta e greu de făcut.

Acesta este primul lui principiu. Îl numeşte „principiul coerenţei structurale”. Acesta presupune coerenţă structurală între „proprietăţile experienţei” şi „proprietăţile proceselor fizice”. Acest lucru pare să se refere la două clase de *fenomene* care, în opinia mea – ca şi a lui –, nu pot fi reduse direct una la cealaltă. Dar Chalmers vorbeşte şi de „tip[uri] de sisteme fizice” şi de „experienţele lor asociate”. Asta pare să se refere la relaţia dintre sistemele *funcţionale* în general şi experienţe în general (prin care cele din urmă „apar din” cele dintâi sau cele dintâi „le generează” pe cele din urmă). Din acest motiv, pentru Chalmers, principiul coerenţei structurale exprimă modul în care „experienţa apare din ce este fizic”.

Din nou, el combină fenomenele fizice şi cauzele „fizice” (adică funcţionale). Dacă nu desluşim etapele intermediare, vom concluziona cu siguranţă că experienţa *nu* apare din ce este fizic. Aceeaşi semnificaţie combinată a „ceea ce este fizic” se poate vedea şi în teoriile funcţionaliste tipice pe care le-am menţionat mai sus. Ele căutau cheia experienţei psihologice în mecanismul *fiziologic* X, Y sau Z:

sincronismul oscilaţiilor gama, fixarea şi stocarea activităţii integrate senzoriale şi frontoparietale, activarea cortexului de către nucleii intralaminari ai talamusului, „ansele de reintrare” talamocorticale etc. Folosind analogia cu tunetul şi fulgerul, această combinare impune să explicăm fenomenele auditive cu ajutorul mecanismelor funcţionale ale vederii. Numai în acest sens are Chalmers dreptate să spună că „o descriere simplă a funcţiilor rămâne de o parte a lacunei, aşa că materialele pentru punte trebuie căutate în altă parte”.

Pentru a explica psihologia în relaţie cu fiziologia, trebuie să ne abstragem din fenomenele observate de *ambele tipuri* (adică trebuie să inferăm mecanismele funcţionale, de ambele tipuri) – apoi să ne abstragem din cele două seturi de abstracţiuni ca să vedem numitorul comun unificator. Când procedăm astfel, trebuie să ne luăm *aceeaşi distanţă* faţă de ambele (adică să inferăm mecanisme suficient de profunde pentru a explica atât funcţiile psihologiei, cât şi pe ale fiziologiei); numai după aceea putem să compatibilizăm fenomenele şi mecanismele lor de bază.

Nu vreau să dau prea multe şabloane. Ştiinţa adevărată evoluează pe căi mai puţin ordonate. Pornim de la o idee sau o descoperire undeva, în cascada cauzalităţii, apoi o completăm cu ce lipseşte. Dar, dacă sărim peste etapele intermediare, ajungem la cerinţa soluţiilor „nonreductive” la problema dificilă care sunt imposibil de obţinut. Să ne amintim primele întrebări ale lui Chalmers: de ce ar trebui ca prelucrarea fizică *să dea naştere* unei vieţi interioare bogate? Cum şi de ce

produc activităţile neurofiziologice experienţa conştiinţei? La aceste întrebări nu se poate răspunde într-un mod nonreductiv, din cauza felului în care sunt formulate.

Prin urmare, este de înţeles concluzia lui Chalmers că o soluţie reductivă, în sine, este imposibilă:

> Sugerez că o teorie a conştiinţei trebuie să considere experienţa fundamentală. Ştim că o teorie a conştiinţei necesită adăugarea a *ceva* fundamental la ontologia noastră, căci totul în teoria fizică este compatibil cu absenţa conştiinţei [...]. O teorie nonreductivă a experienţei va specifica principiile de bază care ne spun cum anume depinde experienţa de caracteristicile fizice ale lumii. Aceste principii *psihofizice* nu vor interfera cu legile fizice, căci se pare că legile fizice deja formează un sistem închis. Mai degrabă, vor fi o completare la o teorie fizică. O teorie fizică dă o teorie a proceselor fizice, iar o teorie psihofizică ne spune cum anume creează procesele respective experienţa. Ştim că experienţa depinde de procesele fizice, dar ştim şi că această dependenţă nu poate fi derivată numai din legile fizice. Noile principii de bază postulate de o teorie nonreductivă ne oferă ingredientul suplimentar de care avem nevoie ca să construim o punte explicativă [...]. Această poziţie se califică drept o variantă de dualism, întrucât postulează proprietăţi de bază ce sunt mai mult decât proprietăţile invocate de fizică. Dar e o variantă inocentă de dualism.

Şi aşa, în cele din urmă, Chalmers spune ceva cam sinistru. Deoarece crede că „experienţa” nu poate fi redusă la „ceea ce e fizic” (în accepţiunea lui, însemnând atât „fiziologic”, cât şi „funcţional”), este obligat să concluzioneze că *experienţa nu face parte din universul fizic cunoscut.* Acesta ar părea să fie dualism pur, de tipul nu-chiar-aşa-de-inocent moştenit de la Descartes şi Locke[32]. Iată de ce Chalmers pretinde că conştiinţa necesită adăugarea a „ceva fundamental la ontologia noastră”, ceva *nonfizic* ce este „mai mult decât proprietăţile invocate de fizică”, ceva care „să completeze” legile fizice etc. Sper că este clar de ce nu sunt de acord cu el în acest sens. Având în vedere tot ce ţi-am spus, nu pot fi de acord că „tot din teoria fizică este compatibil cu absenţa conştiinţei”. Şi mai am ceva de spus. După ce îşi adoptă poziţia împotriva reducerii „experienţei” la „ceea ce e fizic”, Chalmers spune (ca şi cum ar vrea să exemplifice combinarea) că ele sunt, cu toate acestea, aspecte duale a altceva. Acest „altceva”, consideră el, este *informaţia* – care, sunt de acord, nu e ceva fizic în sensul *fiziologic,* dar care, în opinia mea, e ceva fizic în sensul *funcţional* al mecanicii statistice. Această mişcare introduce al doilea principiu al lui Chalmers. El îl numeşte „principiul dublului aspect”, despre care spune că e mai fundamental decât principiul coerenţei structurale:

> Principiul de bază pe care îl sugerez implică în mod esenţial ideea de *informaţie.* Înţeleg informaţia mai mult sau mai puţin în sensul precizat de Shannon (1948). Unde există informaţie, există şi *stări ale*

informaţiei încastrate într-un *spaţiu al informaţiei* […]. Un spaţiu al informaţiei este un obiect abstract, dar, conform lui Shannon, putem înţelege informaţia ca fiind *întruchipată fizic* unde există un spaţiu de stări fizice distincte, diferenţele dintre ele putând fi transmise pe o traiectorie cauzală descendentă […].
Principiul dublului aspect izvorăşte din observaţia că există un izomorfism direct între anumite spaţii ale informaţiei întruchipate fizic şi anumite spaţii ale informaţiei *fenomenologice* (sau experienţiale). Din acelaşi tip de observaţii care au intrat în principiul coerenţei structurale putem vedea că diferenţele dintre stările fenomenologice au o structură ce corespunde direct diferenţelor înglobate în procesele fizice. În special, corespunde diferenţelor care contează semnificativ pe anumite traiectorii cauzale implicate în controlul şi disponibilitatea globale. Adică putem găsi *acelaşi* spaţiu al informaţiei abstracte înglobat în procesarea fizică şi în experienţa conştientă. Asta duce la o ipoteză firească – informaţia (sau cel puţin unele informaţii) are două aspecte de bază, unul fizic şi unul fenomenologic –, ce are statutul unui principiu fundamental care ar putea să susţină şi să explice apariţia experienţei din ceea ce este fizic. Experienţa apare în virtutea statutului său de aspect al informaţiei, când celălalt aspect se găseşte înglobat în procesarea fizică.

De observat că aici Chalmers descrie de două ori informaţia drept ceva „abstract”. Aşadar, odată ce devin clare cele două semnificaţii a „ceea ce e fizic”,

înţelegem că informaţia are acelaşi statut ontologic în raport cu aspectele sale fiziologice şi psihologice ca al electricităţii în raport cu fulgerul şi tunetul. Fulgerul şi tunetul sunt aspecte duale ale descărcării electrice la fel cum semnalele propagate de la retină şi senzaţiile de roşu- aprins sunt aspecte duale ale prelucrării informaţiei – ele sunt răspunsuri diferite evocate de echipament, manifestări fenomenologice diferite ale unui proces cauzal unitar. Iar acest proces cauzal este *fizic* în sensul explicativ.

Chalmers îl citează chiar pe fizicianul John Wheeler în această privinţă. După cum am explicat mai sus, Wheeler a introdus o interpretare „participativă” a mecanicii cuantice, în termenii căreia universul observat se manifestă ca reacţie la întrebările care se pun despre el. Din acest motiv, acelaşi lucru poate lua forme complementare, cum se întâmplă cu undele şi particulele. Interpretarea lui Wheeler este că forma fenomenologică pe care o ia universul trăit depinde de modul în care este observată sau măsurată – adică de perspectiva fiecărui individ asupra lui. De exemplu, o amintire poate fi trăită fie ca o reminiscenţă, fie ca o urmă neuronală activată, în funcţie de echipamentul pe care îl foloseşti ca să o observi. Forma pe care o ia depinde de privitor.

Dar Chalmers o vede diferit. Pentru el, complementaritatea stă în *lucrul observat* şi nu ţine de observator sau de actul observării:

> Legile fizicii pot fi exprimate în termenii informaţiei, postulând stări diferite care generează efecte

diferite, fără să spună de fapt ce *sunt* aceste stări. Contează numai poziţia lor într-un spaţiu al informaţiei. Dacă aşa este, atunci informaţia e un candidat firesc pentru a juca un rol şi într-o teorie fundamentală a conştiinţei. Suntem ghidaţi spre o concepţie a lumii în care informaţia este cu adevărat fundamentală *şi are două aspecte de bază, corespunzând caracteristicilor fizice şi celor fenomenologice ale lumii*[33].

Aşadar, Chalmers crede că aspectele duale reprezentate de „caracteristicile fenomenologice ale lumii” şi „ceea ce este fizic” se află chiar *în* informaţie – la sursa ei –, şi nu în echipamentul observatorului participant[34]. E ca şi cum am crede că dulceaţa este intrinsecă structurii moleculare a glucozei[35]. Poate că această deosebire nu contează când lucrul observat este observatorul, ca atunci când îţi vezi în oglindă propriul corp[36]. Şi poate că, până la urmă, nu contează atât de mult că Chalmers combină două semnificaţii a „ceea ce este fizic” (adică nivelul fenomenologic-fiziologic şi cel mecanic-funcţional), întrucât, cu toate acestea, le reduce pe amândouă la „informaţie”.

Rămânem totuşi cu două probleme. În primul rând, pentru Chalmers, proprietăţile experienţei şi cele a ceea ce este fizic nu sunt *reductibile* la informaţie. Sunt toate *inerente* informaţiei. În al doilea rând, pentru el, toate sunt inerente *tuturor* informaţiilor. Iată motivul pentru care Chalmers afirmă în altă parte a articolului său că „proprietăţile fenomenologice sunt aspectele interne ale informaţiei”, pe care le deosebeşte de proprietăţile

fizice (aşa-zis nonfenomenologice), care sunt „aspectul extern”. Aceasta este trecerea ciudată la care îl duce dualismul său.

Chalmers concluzionează:

> Acest lucru ar putea răspunde la o preocupare despre relevanţa cauzală a experienţei – o preocupare firească, dată fiind o perspectivă în care domeniul fizic este închis din punct de vedere cauzal şi în care experienţa este suplimentară faţă de ceea ce e fizic. Perspectiva informaţională ne permite să înţelegem cum anume ar putea experienţa să aibă un tip subtil de relevanţă cauzală în virtutea statutului său de natură intrinsecă a ceea ce e fizic.

Sper că această carte te-a convins că relevanţa cauzală a experienţei pentru sisteme complexe cu autoorganizare, cum suntem noi, oamenii, nu este în niciun caz subtilă.

Chiar şi aşa, tot trebuie să explicăm de ce şi cum apare experienţa într-un mod justificat din mecanismele fizice. Chalmers întreabă: „De ce îndeplinirea acestor funcţii [fizice] este însoţită de experienţă?”. La fel ca Nagel înaintea lui, el defineşte experienţa în termeni de „ceva-cum-e”. Pentru Chalmers, aceasta este esenţa subiectivităţii, „aspectul intern al informaţiei”, care este, la rândul său, „natura intrinsecă a ceea ce este fizic”. Apoi, afirmă că subiectivitatea nu poate fi redusă la ceva nonsubiectiv. Iată de ce, pentru el, experienţa subiectivă trebuie să fie inerentă informaţiei.

Dar asta pare să implice că calitatea experiată de a fi

ceva (de a fi informaţie) este o proprietate fundamentală a tuturor lucrurilor. De ce crede Chalmers că faptul de a fi *experiat* este „aspectul intern al informaţiei” în general? Dacă afirmăm, cum fac eu aici, că subiectivitatea este doar o perspectivă observaţională, unde un subiect este pur şi simplu faptul de a fi al unui anumit tip de obiect, atunci suntem liberi să spunem că pentru multe obiecte – de fapt, pentru majoritatea – *nu* există „ceva cum e” să fii ele. *Majoritatea covârşitoare a lucrurilor nu au acel ceva-cum-e* când sunt analizate din propria perspectivă subiectivă.

Există ceva cum e să fii un organism unicelular sau o plantă? Dar o piatră? Există ceva cum e să fii un calculator sau internetul? Dar un termostat? Chalmers spune: „E indiscutabil că unele organisme sunt subiecte ale experienţei”. Înseamnă asta că *numai* organismele (sau numai *anumite* organisme) sunt subiecte ale experienţei? El recunoaşte că acestea sunt întrebări deschise şi că se poate ca experienţa să apară numai la un anumit nivel de complexitate sau cu un anumit tip de prelucrare a informaţiei[37]:

> O întrebare evidentă este dacă *toate* informaţiile au un aspect fenomenologic. O posibilitate este că avem nevoie de încă o constrângere asupra teoriei fundamentale, care să indice exact *tipul* de informaţie ce are un aspect fenomenologic. Cealaltă posibilitate este că nu există o asemenea constrângere. Dacă nu, atunci experienţa este mult mai răspândită decât am crezut, întrucât informaţia

este pretutindeni. Ceea ce este ilogic iniţial, dar, dacă mă gândesc mai bine, cred că această poziţie capătă o anumită plauzibilitate şi eleganţă. Acolo unde există prelucrare simplă a informaţiei, există experienţă simplă, iar acolo unde există prelucrare complexă a informaţiei, există experienţă complexă. Un şoarece are o structură de prelucrare a informaţiei mai simplă decât un om şi, în consecinţă, o experienţă mai simplă; poate că un termostat, o structură de prelucrare a informaţiei maxim de simplă, are o experienţă maxim de simplă? Într-adevăr, dacă experienţa este într-adevăr o proprietate fundamentală, ar fi surprinzător să apară doar din când în când; majoritatea proprietăţilor fundamentale sunt răspândite mai uniform. În orice caz, aceasta rămâne în mare măsură o întrebare deschisă.

Dacă accepţi opinia lui Chalmers că experienţa este o proprietate fundamentală a orice, atunci nu e nimic de explicat; de asemenea, dacă experienţa este pretutindeni şi eternă, atunci poate că nu avem de ce să ne temem. Cu preţul unei singure supoziţii absurde, ignorăm problema dificilă şi ne acordăm nemurirea, debarasându-ne de temerea că existenţa noastră ar putea *depinde* de ceva ca şi cum ar fi doar un coşmar.

De-ar fi aşa. Din păcate, după cum am văzut în capitolele anterioare, experienţa apare numai în condiţii speciale. Ea se află într-un loc precis printre fluxurile de informaţii care trec prin fiinţa noastră jalnic de fragilă. Îndeplineşte anumite sarcini, apoi dispare. După câte se pare, nu e cu nimic mai fundamentală pentru structura

realităţii decât orice alt tip de răspuns evocat de echipament. Îmi pare foarte rău. Dacă nu accepţi speculaţia lui Chalmers, atunci trebuie să răspundem la întrebarea: de ce există ceva cum e în cazul unor lucruri, dar nu şi al altora? Dacă nu toate obiectele sunt subiecte şi dacă nu toate subiectele sunt capabile să simtă, atunci cum se întâmplă că *uneori* apar subiecte capabile să simtă?

Mă contrariază propunerea expansivă a lui Chalmers despre informaţie, dar cred că este mult mai plauzibil că *„avem nevoie de încă o constrângere* asupra teoriei fundamentale, care să indice exact *tipul* de informaţie ce are un aspect fenomenologic”. Definirea acestei constrângeri este sarcina principală a volumului de faţă. Am văzut că totul se învârte în jurul scopului şi obiectivului prelucrării informaţiei. Când vorbim de conştiinţă, asta înseamnă *minimizarea entropiei.* Dar e mai complicat: avem nevoie şi de o zonă Markov. Apoi, presupune minimizarea *propriei tale* entropii. Mai mult, presupune acest lucru la nivelul foarte multor parametri categoriali în contexte neprevăzute. Realităţile fizice pe care le- am evidenţiat în această carte relevă că conştiinţa nu este intrinsecă tuturor informaţiilor, ci unui anumit tip de prelucrare a informaţiilor: o formă complexă a tipului celor care se autovalidează.

Dacă numai unele lucruri sau dacă numai organismele sau dacă numai unele organisme sunt subiecte ale experienţei, atunci conştiinţa nu poate fi o proprietate fundamentală a universului. Cu siguranţă că a existat un început al vieţii: există numeroase dovezi

empirice în acest sens. Iar viaţa a apărut la mult, mult timp după Big Bang. Aşadar, cu siguranţă a trecut mult timp înainte să existe conştiinţă. Doar această presupunere – că a existat un început al conştiinţei – ne obligă să-i găsim o explicaţie fizică. Dacă a existat un început, trebuie să fi fost ceva înainte de conştiinţă care o explică. Ideea alternativă – conştiinţa a precedat viaţa şi universul – nu se potriveşte faptelor aşa cum par să fie şi, mai mult, sună în mod inutil ca ideea de Dumnezeu. Dacă nu invoci asemenea idei, conştiinţa trebuie să fi apărut – prin urmare trebuie să fie o parte – dintr-un univers fizic nonconştient.

Această afirmaţie este valabilă chiar în interiorul nostru. Am văzut în capitolele anterioare că percepţia şi cogniţia nu sunt cu necesitate însoţite de conştiinţă. De fapt, dovezile ştiinţifice sugerează că percepţia şi cogniţia sunt *în general* inconştiente. În legătură cu asta am citat articolele analitice clasice de Kihlstrom şi de Bargh şi Chartrand. După ce au analizat dovezile, aceşti specialişti au ajuns la concluzia că, „în cea mai mare parte a timpului”, nu suntem conştienţi de actele noastre psihologice.

Multe dintre dovezile empirice analizate de aceşti autori merită atenţia noastră (mai ales modul în care intenţionalitatea inconştientă derivă din învăţarea conştientă din experienţă[38]), dar vreau să amintesc numai ideea centrală: au loc şi există evenimente în interiorul tău indiferent dacă eşti conştient de ele sau nu, inclusiv evenimente *psihologice.* Am fi putut să ajungem la această concluzie, într-un anumit sens evidentă, plecând de la multe premise diferite, dar asta

nu contează. Dacă cea mai mare parte a „vieţii psihice din fiecare clipă” continuă fără experienţă, atunci de ce nu se poate ca termostatele şi internetul să fie inconştiente? Dacă existenţa ganglionilor bazali nondeclarativi din creier nu este „intrinsec experienţială” (ca să folosesc termenul lui Chalmers), atunci cum pot orice informaţii să fie experienţiale?

În mod clar, subiectivitatea în general trebuie să fie alcătuită din ceva nonexperienţial. Aceasta nu este o idee metafizică despre lucrurile-în-sine versus lucrurile-aşa-cum-le-experiem, ci un fapt empiric. Viaţa psihică din fiecare clipă *conform experienţei mele subiective* nu e în niciun caz totalitatea vieţii mele psihologice. De aici se vede importanţa experienţei. După cum ne-a învăţat Freud, nu orice din viaţa psihologică este conştient. Şi putem să înţelegem natura conştiinţei doar dacă ne dăm seama ce face ea mai exact. Adică ce funcţie are.

Mă tot întreb: ce *adaugă* conştientizarea la prelucrarea informaţiilor – la prelucrarea informaţiilor care se întâmplă oricum, în mod inconştient? Ce rost are să devii conştient de procesele fizice, dacă conştientizarea ta nu are nicio influenţă asupra lor? Mi se pare că această întrebare coincide cu întrebarea esenţială a lui Chalmers: „De ce îndeplinirea acestor funcţii este însoţită de experienţă? [...] De ce toată această prelucrare a informaţiilor nu continuă «pe întuneric», fără să se simtă nimic în interior?”.

În opinia mea, întrebarea a apărut numai pentru că Chalmers, ca şi Crick, a căutat funcţia conştiinţei în locul greşit. Forma fundamentală a conştiinţei nu e ceva

cognitiv, ca vederea, ci e ceva afectiv. Chalmers a avut dreptate – dar numai în acest sens – să sugereze că conştiinţa nu este o funcţie *cognitivă:* funcţia primară a conştiinţei nu este să perceapă, să- şi amintească sau să înţeleagă, ci *să simtă.*

Cum ar putea funcţia de a simţi să continue „pe întuneric", fără niciun sentiment? Suntem îndreptăţiţi să întrebăm de ce vederea este însoţită de experienţă. Vederea nu necesită conştiinţă şi nici oricare dintre celelalte procese cognitive. Dar simţirea da[39]. E adevărat că unii teoreticieni afirmă că pulsiunile, afectele şi emoţiile nu sunt *cu necesitate* conştiente. Asta pentru că oameni diferiţi se referă la lucruri diferite prin aceste cuvinte. Iată de ce folosesc cuvântul „sentiment": dacă există emoţii inconştiente, „sentiment" se referă la corespondentele lor care sunt conştiente. Nu poate să existe sentiment pe care nu îl simţi. „Sentiment inconştient" este un oximoron.

Acesta este motivul pentru care am centrat argumentele ştiinţifice din cartea de faţă pe sentiment. *Ca să rezolve problema dificilă a conştiinţei, ştiinţa trebuie să identifice legile care guvernează funcţia psihică a „simţirii".* Nu e vorba doar de felul cum îmi aleg cuvintele. Am prezentat dovezi semnificative ca să arăt că sentimentul este forma fundamentală a conştiinţei, premisa ei obligatorie. De asemenea, am explicat atât din punct de vedere fiziologic, cât şi mecanicist diferenţa dintre nevoile resimţite şi cele neresimţite şi am arătat că sentimentele au consecinţe concrete. Asta mi-a permis să concluzionez în capitolul 9 că „conştiinţa nu este doar o perspectivă subiectivă

asupra dinamicii «reale» a sistemelor cu autoorganizare, ci şi o funcţie având propriile puteri cauzale determinate”.

Nedumerirea din întrebările lui Chalmers – „De ce îndeplinirea acestor funcţii este însoţită de experienţă?” şi „De ce toată această prelucrare a informaţiei nu continuă «pe întuneric»?” – dispare când ne dăm seama că „experienţa” nu este intrinsecă vederii şi prelucrării informaţiei în general. Ea e intrinsecă numai formei specifice de prelucrare a informaţiei care generează sentiment.

Iată de ce consider eu că *nu* există ceva cum e să fii termostat sau internetul. Aceasta este pertinenţa crucială a întrebărilor „De ce există ceva cum e în cazul unor lucruri, dar nu şi al altora?” şi „De ce (şi cum) se întâmplă că *uneori* apar subiecte capabile să simtă?”.

La aceste întrebări trebuie să se răspundă în termeni mecanicişti. Care este funcţia sentimentului? Poate sentimentul să-şi îndeplinească funcţia în lipsa experienţei? Dacă Mary, o persoană oarbă, specialistă în corelativele neurale ale vederii, ar fi specialistă în corelativele neurale ale *afectului,* oare cunoştinţele ei legate de tot ce se poate şti despre funcţia de a simţi nu ar explica de ce (sau chiar nu ar prezice că) conştientizezi că se simte cumva?[40] La fel şi în cazul spectrelor inversate ale lui Locke. Este oare posibil din punct de vedere logic să simţi ca înfiorător de dureros ceva ce eu simt ca extraordinar de delicios, iar asta să nu aibă nicio importanţă? Sigur că nu. Şi asta pentru că senzaţiile chiar *fac* ceva – şi, pe parcurs, ne cresc

semnificativ şansele să supravieţuim[41].

E uşor de recunoscut lacuna explicativă dintre prelucrarea informaţiilor vizuale şi conştientizarea vizuală. Dar nu există acelaşi tip de lacună între funcţia de a simţi şi experienţa simţirii ei. Unii filosofi vor spune că există în continuare o asemenea lacună. Vor arăta că „zombii afectivi" sunt *imaginabili*[42]. Aş zice că acest lucru se va baza pe preocupările istorice ale disciplinei lor, şi nu pe o nedumerire rezonabilă, aşa că fac un apel la ei să analizeze chestiunea dintr-o perspectivă nouă.

Preocupările legate de „lacuna explicativă" nu ar fi apărut niciodată dacă ne-am fi început căutarea întrebând de ce şi cum apar sentimentele, şi nu încercând să găsim un corelativ neural al conştiinţei în cortexul vizual. Funcţia biologică a senzaţiilor ca foamea nu e nimic misterios, iar acel ceva-cum-e al lor nu este deosebit de dificil de explicat. Trebuie pur şi simplu să urmezi logica minimizării energiei libere acolo unde duce în cazul sistemelor cu autoorganizare, cum suntem noi. Date fiind numeroasele noastre nevoi, mediile complexe şi periculoase, gama largă de acţiuni posibile şi capacitatea de a executa doar una sau două dintre ele în acelaşi timp, ar trebui să ne *aşteptăm* să avem o lume interioară, construită în scopul deliberării şi alegerii. Şi cu ce ar trebui să ne aşteptăm să fie umplută? Cu ce altceva dacă nu cu o diversitate dinamică de calităţi de evaluare, incluzând, cu rol central, cântăriri ale încrederii, ce urmăresc şi măsoară diversele noastre nevoi incomensurabile pe măsură ce

apar, alături de caracteristicile predominante ale mediului în care trebuie satisfăcute.

Gândeşte-te la următoarea afirmaţie a lui Chalmers. Ar mai fi făcut el asemenea afirmaţii dacă ar fi vorbit despre funcţii afective, şi nu cognitive?

> Prin asta nu vreau să spun că experienţa *are* zero funcţii. Poate se va dovedi că joacă un rol cognitiv important. Dar, orice rol ar putea să joace, explicarea experienţei va presupune mai mult decât o simplă explicare a funcţiei sale. Poate se va dovedi chiar că, în cursul explicării unei funcţii, vom ajunge la ideea esenţială ce permite o explicare a experienţei. Totuşi, dacă se întâmplă aşa, această descoperire va fi o recompensă explicativă *suplimentară.* Nu există o funcţie cognitivă astfel încât putem spune dinainte că explicarea acestei funcţii va explica *automat* experienţa[43].

Eu pot spune dinainte că explicarea funcţiei de a simţi va explica automat experienţa. Nu văd cum o descriere adecvată a ei, din perspectiva ştiinţelor naturale, ar putea să rateze acest lucru. Începând de la termodinamică, ajungem – surprinzător de uşor – la o subiectivitate calificată şi manifestând intenţionalitate, ale cărei priorităţi de maximă urgenţă sunt cântărite pentru o clipă, simţite şi apoi transformate cu (sperăm) circumspecţia cuvenită în acţiune în derulare. Asta, aş spune, este cum e să fii orice în legătură cu care există ceva cum e să fii respectivul lucru.

Lămureşte asta şi ultimul mister despre natura conştiinţei? Aici trebuie să recunosc că mai e ceva care

nu-mi dă pace. Ce ciudat ar fi dacă tot ce s-a trăit vreodată – dacă însăşi cognoscibilitatea universului – ar depinde de mecanismele pe care le-am descris. Gândul la aşa ceva e destul să-mi revolte sensibilitatea. Totodată, e ciudat până şi că existăm, noi sau orice altceva. De realitatea contingenţei nu te poţi ascunde. Acest lucru poate fi tulburător chiar şi în contexte mai obişnuite – ca atunci când scapi ca prin urechile acului, când îţi dai seama că, dacă nu ar fi existat o oarecare situaţie nebănuită, poate că nu ai fi supravieţuit. Şi poate chiar *ar trebui* să ne tulbure: dacă privim lucrurile dintr-o anumită perspectivă, sentimentele noastre au evoluat tocmai ca să nu ajungem să vedem moartea cu ochii. Îmi amintesc din nou de domnul S., a cărui subiectivitate muncea din greu să ascundă precaritatea existenţei lui după intervenţia chirurgicală nereuşită. Şi nu e singurul. Dacă ni se oferă o descriere a fiinţei noastre care o înfăţişează ca fiind la fel de fragilă ca a mea, poate că un impuls spre negarea plină de iluzii e pur şi simplu firesc.

Dar e mai mult decât atât. Toate explicaţiile trebuie să ia ceva drept dat şi, prin urmare, inexplicabil în cadrul teoriei. Orice poveste trebuie să se termine undeva. Pentru mine, drumul se încheie cu informaţia, fără îndoială ceva deconcertant, şi cu autoorganizarea, indiscutabil misterioasă. În felul cum prezint eu conştiinţa, totul izvorăşte din pulsiunea sistemului de a exista. Minţile noastre sunt urzite din ordinea însăşi, ce apare spontan din haos ca în experimentul lui Friston, apoi se apără de atacurile entropiei. Cum poate asta să fie baza existenţei noastre? Ce *este* ordinea, încât are

asemenea puteri să ne facă să apărem din întunericul neînsufleţit de dinaintea noastră şi de dincolo de noi? Aceste întrebări depăşesc sfera cărţii de faţă. Din câte ştiu, ele scapă cu totul cercetării. Însă miar plăcea să le ştiu răspunsul.

Cu excepţia acestei revelaţii, sper că ceea ce ofer eu este totuşi valoros. Până la urmă, e ceva despre care mulţi s-au îndoit că ar putea fi posibil vreodată: o descriere a capacităţii de a simţi prin prisma altor lucruri despre care ştim că există în lumea fizică. Rămân încă multe probleme dificile, dar poate că nu problema dificilă – oricum, nu chiar în forma cu care ne-am obişnuit. Şi, dacă uneori până şi eu am îndoieli că conştiinţa poate fi ceea ce spun eu că este – dacă nu mă simt 100% sigur de descrierea mea –, mă consolez cu faptul că incertitudinea supărătoare, fără perspectiva unei clarificări finale, este exact ce ar prezice teoria mea.

*1 Îmi amintesc bine experienţa primei operaţii pe creier la care am asistat. Pe o parte a capului pacientului părul a fost ras şi s-a aplicat un lichid maroniu-gălbui. Apoi, pe scalp a fost trasată o curbă amplă cu un marker. Pielea a fost tăiată cu un bisturiu mânuit cu dexteritate de-a lungul liniei respective, apoi a fost pliată spre partea din spate a capului, dezvelind craniul. Pe suprafaţa rotunjită s-au prelins nişte firicele de sânge. Au fost şterse cu calm. Apoi, cu un burghiu s-au realizat, manual, patru găuri mari în os. M-am temut că vârful burghiului va pătrunde în creier. În continuare, o lamă subţire de

fierăstrău a fost introdusă prin găuri şi trasă dintr-o parte în alta. Mirosea a ars. Această procedură a fost repetată de patru ori, apoi porţiunea de os astfel decupată a fost scoasă şi pusă într-un vas. A urmat dura mater, o membrană groasă care stătea în cale şi care a sângerat abundent în timp ce era tăiată. Fiecare vas de sânge a fost cauterizat cu răbdare. Procedura a ajuns la punctul culminant. Dura mater a fost dată la o parte ca un văl, apoi le-am văzut: circumvoluţiunile de un roz pal ale cortexului însuşi, cu nişte vase de sânge şerpuite cuibărindu-i-se în şanţuri. Am fost copleşit de un sentiment de *evlavie.* Parcă intram într-o catedrală. A urmat sarcina pe care aveam să o îndeplinesc şi eu într-o zi: o serie scurtă de conversaţii focusate cu pacientul, pe măsură ce un electrod îi intra printre circumvoluţiunile gelatinoase şi lucioase ale creierului, înainte să se facă incizia aferentă – o incizie în minte…

12

Cum se face o minte

În copilărie, mi se părea că nu are rost să fac nimic. Indiferent ce aş fi făcut, tot urma să dispar pentru totdeauna. Conştiinţa mea era inseparabilă de creierul meu şi, judecând după toate dovezile, şi asta era ceva strict limitat în timp. Acest lucru mă tulbura profund. Singura cale de ieşire din groapa nihilismului pe care am putut să o găsesc a fost să încerc să înţeleg ce *este* conştiinţa. Dacă eram sincer şi depuneam eforturi concertate în acest demers, măcar nu îmi iroseam scurta existenţă, atâta cât mi se alocase. Aş fi investit-o în singura problemă ce merita să fie rezolvată în circumstanţele respective. Dacă procedam astfel, aveam şi speranţa – slabă, dar nu irealizabilă – că, dacă aş înţelege ce e conştiinţa, poate aş reuşi să-i evit limitele. Poate aş găsi cumva o cale de a scăpa din bula solipsistă a existenţei, de a contextualiza „faptul de a fi” în cadrul unei perspective de ansamblu. Mărturisesc că, prin asta, speram că aş putea găsi chiar vreo alternativă la înspăimântătoarea implicaţie logică a mortalităţii.

Aşa că am ales calea ce a culminat cu cartea de faţă. Pe parcurs – în timpul formării mele în neuropsihologie

şi în psihanaliză –, m-am consolat cu descoperirea că logica este produsul unui instrument limitat. Am dobândit o cunoaştere directă a câtă „viaţă psihică din fiecare clipă” apare în afara conştientizării şi a controlului voluntar. Adică am înţeles locul gândirii. Am văzut şi cum unii pacienţi cu antecedente neurologice nu erau conştienţi nici de unele dintre adevărurile cele mai evidente, din cauză că îşi pierduseră anumite părţi ale creierului. Dacă oameni ca domnul S. sunt împinşi spre asumpţii atât de incorecte de anumite limitări ale instrumentelor lor psihice, poate că acelaşi lucru e valabil şi în cazul meu. Dar dacă nouă tuturor, ca şi lor, ne lipseşte maşinăria care, dacă am avea-o, ne-ar permite să ajungem la concluzii radical mai bune despre noi înşine şi despre locul nostru în univers? Dacă toţi am avea numai simţul auzului, poate că am crede că realitatea constă din ceva eteric precum undele sonore şi nu am avea nicio idee despre lumea vizibilă şi tangibilă a corpurilor solide psihice. Întrucât toţi am fi constrânşi de aceleaşi dovezi incomplete, toţi am ajunge la aceleaşi concluzii greşite. Conform aceleiaşi logici, dacă am avea nişte creier în plus – cinci lobi per emisferă, în loc de cei patru obişnuiţi, să zicem –, poate că am şti ceva despre natura lucrurilor care acum ne scapă.

Poate. Dar e la fel de adevărat şi că faptele necunoscute s-ar putea dovedi a fi *mai* deprimante decât cele cu care ne chinuim în prezent. Unii ar putea să afirme că, pe cât eşti mai inteligent, pe atât mai rău – nu mai bun –, faţă de cât credeai, se dovedeşte a fi locul tău în univers. Nu sunt sigur că acest punct de vedere

se justifică. În orice caz, nu putem decât să ne folosim resursele cât mai bine. Regulile ştiinţei cer să ne testăm conjecturile cu cele mai bune dovezi pe care le găsim, fiind pregătiţi să respingem ipotezele infirmate de dovezi. Regulile vieţii cer acelaşi lucru de la experienţa de zi cu zi. După cum stau lucrurile, există insuficiente dovezi care să susţină ipoteza că fiinţa mea capabilă să simtă va supravieţui corpului meu muritor. Se pare că suntem cu totul dependenţi de creierul nostru alarmant de fragil.

Pornind de la această premisă, putem presupune că conştiinţa nu a existat pe pământ înainte să se dezvolte creierul – şi poate că abia când s-a dezvoltat creierul vertebratelor; deci până acum aproximativ 525 de milioane de ani. Presupun că, mai înainte, a apărut într-o formă rudimentară, că un precursor al afectului a devenit treptat afect resimţit, fără să existe demarcaţii clare între ele, în tandem cu evoluţia unor organisme tot mai complexe, având multe nevoi concurente. Odată cu evoluţia cortexului a apărut conştiinţa *cognitivă* – adică abilitatea suplimentară de a contextualiza afectul în mod exteroceptiv şi de a-l păstra în minte[1]. În orice caz, fiinţa capabilă de simţire nu are cum să fi existat înainte să existe sisteme nervoase. Forma subiectivă, interioară a conştiinţei nu poate să fi existat în lipsa corpului ei exterior. Prin urmare, în baza dovezilor pe care le-am prezentat în capitolul 6, putem concluziona că conştiinţa aşa cum o ştim presupune existenţa a ceva care arată ca PAG sau ca precursorul său evolutiv imediat, precum şi a echipamentului adiacent din triunghiul de decizie din mezencefal şi din sistemul

reticulat activator.

În timp ce citeşti aceste cuvinte, poate că în tine apare o urmă de îndoială. Cu siguranţă, am conştientizat-o şi eu cândva, între februarie şi iulie 2018. Începusem să mă întâlnesc în mod regulat cu un mic grup de fizicieni şi informaticieni care îmi împărtăşeau sau cel puţin îmi înţelegeau opinia că redefinirea conştiinţei ca sentiment ar putea deschide calea explicării ei
dintr-o perspectivă fizică[2]. Pe măsură ce am avansat cu schimburile noastre de idei, mi-am dat seama că adoptam o perspectivă care poate că nu te va surprinde acum, după ce ai citit această carte – dar iniţial pe mine m-a surprins. Gândul tulburător care mi-a venit era: conştiinţa aşa cum o ştim din [illegible] interior nu presupune cu necesitate existenţa a ceva care *arată* ca PAG, ci doar existenţa a ceva care *funcţionează* ca PAG.

David Chalmers, spre deosebire de mulţi dintre discipolii săi, crede că problema dificilă poate fi rezolvată. De fapt, în acelaşi articol în care a prezentat iniţial problema, a schiţat şi o soluţie posibilă pentru ea cu titlul: „Outline of a theory of consciousness” (Schiţa unei teorii a conştiinţei). Teoria sa nu este foarte bine cunoscută, probabil pentru că puţini specialişti o acceptă. Ea se bazează pe trei principii, iar două dintre ele le- am prezentat în capitolul anterior. Acestea sunt principiul coerenţei structurale şi principiul dublului aspect. Le-am contestat pe amândouă, dar am avut şi o atitudine deschisă. Într-un fel, această carte pur şi

simplu le nuanţează. Am spus că voi prezenta al treilea principiu acum, în ultimul capitol. E vorba de principiul invarianţei organizatorice.

Acest principiu afirmă pur şi simplu că oricare două sisteme cu aceeaşi organizare funcţională rafinată vor avea experienţe identice din punct de vedere calitativ. De exemplu, dacă tiparele cauzale ale organizării neuronale ar fi reproduse în silicon, cu un cip de silicon pentru fiecare neuron şi aceleaşi tipare de interacţiune, atunci ar apărea aceleaşi experienţe. Conform acestui principiu, pentru apariţia experienţei contează nu alcătuirea fizică specifică a unui sistem, ci *tiparul funcţional al interacţiunii cauzale dintre componentele sale.* Desigur, acest principiu este controversat. John Searle, printre alţii, afirmă că conştiinţa este legată de o biologie specifică, astfel încât o izomorfă din silicon a unui om nu ar fi conştientă[3]. Totuşi, eu cred că principiul se poate susţine semnificativ prin analiza experimentelor mentale.

Chalmers schiţează un asemenea experiment mental[4], centrat pe ideea a două sisteme de prelucrare a informaţiilor având organizări funcţionale identice. În unul dintre ele organizarea este produsă de o configuraţie de neuroni (ca în creierul natural), iar în celălalt e produsă de o configuraţie de cipuri din silicon (cum ar putea fi într-un creier artificial). După cum ştim, creierul uman natural este capabil de experienţă. Întrebarea pe care o pune Chalmers este dacă aceeaşi afirmaţie este valabilă şi în cazul unei replici funcţionale exacte a creierului uman. Iată experimentul mental:

> Cele două sisteme au aceeaşi organizare, aşa că ne putem imagina cum se transformă treptat unul în celălalt, poate prin înlocuirea neuronilor, unul câte unul, de către cipuri din silicon având aceeaşi funcţie locală. Astfel, obţinem un spectru de cazuri intermediare, fiecare cu aceeaşi organizare, dar cu o alcătuire fizică uşor diferită[5].

Întrebarea ce apare aici este dacă înlocuirea unui neuron natural cu unul artificial va schimba în vreun fel experienţa creierului natural şi invers. Argumentul ce decurge de aici este că creierul neuronal va continua să aibă aceleaşi experienţe pe măsură ce trece prin toate micile etape intermediare, până şi inclusiv în punctul în care devine o copie integrală din silicon a creierului. (Şi invers, desigur.) Asta demonstrează – spune

Chalmers – că şi creierii din silicon sunt capabili să aibă experienţe:

> Dată fiind asumpţia extrem de plauzibilă că schimbările de la nivelul experienţei corespund schimbărilor de la nivelul prelucrării informaţiilor, suntem conduşi spre concluzia [... întrucât nu a apărut nicio schimbare în procesare în timpul tranziţiei de la neuron la silicon] că *oricare două sisteme izomorfe din punct de vedere funcţional trebuie să aibă acelaşi tip de experienţe*[6].

Chalmers admite că „unii s-ar putea îngrijora că o

izomorfă de silicon a unui sistem neuronal ar putea fi imposibilă din motive tehnice. Această întrebare rămâne deschisă. Principiul invarianţei spune numai că, *dacă* o izomorfă este posibilă, atunci va avea acelaşi tip de experienţă conştientă". Şi concluzionează:

> Acest experiment mental porneşte de la realităţi familiare despre coerenţa dintre conştiinţă şi procesarea cognitivă, pentru a produce o concluzie puternică despre relaţia dintre structura fizică şi experienţă. Dacă argumentul este acceptat, ştim că singurele proprietăţi fizice direct relevante pentru apariţia experienţei sunt proprietăţile *organizatorice*[7].

Eu nu găsesc nicio fisură în această logică. Ea corespunde în linii mari afirmaţiei mele că o organizare funcţională unică (de exemplu, Mark Solms) îşi poate asuma două apariţii diferite, una dintr-o perspectivă introspectivă şi cealaltă dintr-o perspectivă extrospectivă. Nu sunt de acord cu Chalmers numai atunci când foloseşte acest argument pentru a elabora teoria că *toate* informaţiile au un aspect subiectiv – fiind, prin urmare, conştiente. Am spus (citându-l pe Chalmers) că „avem nevoie de încă o constrângere asupra teoriei fundamentale, care să indice exact *tipul* de informaţie ce are un aspect fenomenologic". Doar aici nu suntem de acord. În cartea de faţă am identificat tipul special de prelucrare a informaţiilor despre care cred că au un aspect fenomenologic, explicând de ce şi cum se întâmplă acest lucru. Indiferent dacă Chalmers

acceptă sau nu această versiune „constrânsă” a principiului dublului aspect şi a principiului coerenţei structurale[8], tot suntem de acord că oricare două sisteme izomorfe din punct de vedere funcţional trebuie să aibă acelaşi tip de experienţe. Cu alte cuvinte, dacă unul din două sisteme izomorfe funcţional are experienţe fenomenologice, atunci *şi celălalt trebuie să le aibă.*

Înainte să analizăm implicaţiile acestei concluzii, să ne gândim la îngrijorarea ştiinţifică potrivit căreia o izomorfă din silicon a unui sistem neuronal ar putea fi imposibilă din motive tehnice. Chalmers spune că „rămâne o întrebare deschisă”. Aceste cuvinte au fost scrise în 1995. Întrebarea nu mai este deschisă. Într-o serie de studii publicate între 2012 şi 2016, mai multe grupuri de cercetători au arătat că este posibil să se creeze o interfaţă artificială între creier şi măduva spinării care permite animalelor paralizate să-şi mişte membrele afectate prin înlocuirea neurotransmisiei spinale cu semnale radio[9]. De exemplu, grupul lui Marco Capogrosso a raportat următoarele:

> Unor maimuţe rhesus *(Macaca mulatta)* li s-au implantat o reţea intracorticală de microelectrozi în zona cortexului motor responsabilă de membrele inferioare şi un sistem de stimulare a măduvei spinării compus dintr-un implant epidural selectiv sub aspect spaţial şi un generator de impulsuri cu posibilităţi de declanşare în timp real. Am proiectat şi am implementat sisteme de control wireless care

au conectat online decodificarea neuronală a stărilor motoare de extensie şi de flexie cu protocoale de stimulare ce încurajează aceste mişcări. Aceste sisteme au permis maimuţelor să se comporte liber, fără restricţii sau elemente electronice ataşate constrângătoare[10].

Aceste maimuţe aveau leziuni (unilaterale) la nivelul neuronilor motori superiori din tractul corticospinal de la nivelul coloanei vertebrale toracale, ducând la paralizia unui membru inferior. „La doar şase zile după traumatism şi fără dresajul anterior al maimuţelor, interfaţa creier-măduvă a restabilit locomoţia pentru susţinerea greutăţii membrului inferior paralizat pe o bandă de alergare şi la sol.”[11] Ideea esenţială din acest studiu este că „interfaţa creier-măduvă” (adică semnalele radio care au restabilit comunicarea între cortex şi coloana lombară, ocolind leziunea corticospinală) nu este decât *o izomorfă artificială* de tipul celor de care Searle era îngrijorat că ar putea fi „imposibile din motive tehnice”. Semnalele radio au înlocuit semnalele neuronale exact în modul în care Chalmers a preconizat că cipurile sale imaginare din silicon ar putea înlocui neuronii. În studiul lui Capogrosso, semnalele radio au îndeplinit aceeaşi funcţie pe care o îndeplinesc de obicei neuronii corticospinali.

E adevărat că vorbesc despre neuroni corticospinali care îndeplinesc numai funcţii *motoare* (şi nu cognitive), dar e important de observat că neuronii respectivi sunt

„piramidali”, cu originea în stratul V al neocortexului. Aşadar, sunt identici sub aspect morfologic cu tipul de neuroni descrişi pe larg în literatura despre codificarea predictivă drept neuroni ce procesează atât semnalele „de predicţie”, cât şi semnalele „de eroare” în interferenţa perceptivă, precum şi în cea activă. Neuroni de acest tip se întâlnesc în tot cortexul, inclusiv în hipocamp şi în lobii prefrontali. Prin urmare, ideea fundamentală este: funcţia unui neuron piramidal poate fi îndeplinită de o izomorfă artificială (în acest caz, funcţia lui de interfaţă între neocortex şi muşchi poate fi îndeplinită de undele radio). Şi, ca să se înţeleagă clar acest lucru, nu cu mult timp în urmă a fost creată o izomorfă de silicon complet funcţională a unui neuron piramidal[12].

Acum aş vrea să-ţi îndrepţi atenţia spre aspectul esenţial al studiului pe care tocmai l-am descris. Reţeaua intracorticală implantată în cortexul motor al maimuţelor rhesus a înregistrat *informaţie.* Informaţia produsă de activitatea corticală a fost un *mesaj* pe care creierul maimuţei a încercat să-l trimită membrelor inferioare ale maimuţei, folosind neuroni piramidali corticospinali. Reţeaua de microelectrozi a codificat acest mesaj în unde radio, apoi a transmis aceeaşi informaţie printr-un mediu artificial spre coloana lombară, care apoi a transferat-o înapoi în mediul ei natural, ca să genereze mişcarea intenţionată. Cu alte cuvinte, funcţia îndeplinită artificial de undele radio a fost una de prelucrare a informaţiei care, de regulă, este îndeplinită în mod natural de neuroni piramidali

corticospinali. Acelaşi principiu se aplică neuronilor din silicon construiţi recent, deşi la momentul scrierii acestei cărţi încă nu fuseseră folosiţi în asemenea experimente.

Ca să dezvolt această idee, voi arăta că acelaşi lucru se poate realiza şi în cazul altor funcţii cerebrale de prelucrare a informaţiei, inclusiv funcţiile cognitive specific umane. Nu spun doar că se poate realiza, ci că *s-a şi realizat.* Într-un studiu de referinţă publicat în 2012, Brian Pasley şi colegii săi au înregistrat activitate electrică în reţele intracorticale plasate pe cortexul auditiv în cazurile a 15 voluntari umani, în timp ce aceştia erau supuşi unor intervenţii chirurgicale pe creier din motive medicale[13]. În timpul intervenţiilor, cercetătorii au transmis cuvinte prin difuzoare sau căşti, pe care pacienţii le- au ascultat în mod conştient. Echipa a studiat înregistrările rezultate din EEG-uri şi a construit un algoritm – un model computaţional –, pentru a reprezenta pe tiparele EEG sunetele auzite de pacienţi. Modelul a potrivit fiecare sunet rostit cu tiparul aferent de activitate cerebrală. Apoi, cercetătorii au proiectat invers procesul: pornind de la tiparele de activitate cerebrală (adică informaţia), au arătat că este posibil să foloseşti modelul computaţional pentru a reconstrui cuvintele auzite de creier.

Aici, algoritmul calculatorului joacă un rol echivalent celui îndeplinit de undele radio în experimentul anterior, dar unul mai complex. El modelează *artificial* exact aceeaşi informaţie care este modelată în mod natural de creier, apoi generează sunetul corespunzător

generat de modelul respectiv.

Ce au făcut aceşti specialişti cu cuvintele se poate face – şi sa şi făcut – şi cu imaginile vizuale. Shinji Nishimoto şi Jack Gallant au arătat că se pot dezvolta algoritmi pe calculator care decodifică – doar din imagistică RMNf – ceea ce vede cortexul vizual, transformând ulterior respectivele tipare de activitate cerebrală în imagini în mişcare[14]. În acest mod, pornind doar de la imagistică RMNf, se pot genera aproximări rezonabile ale imaginilor vizuale care au cauzat activitatea cerebrală.

Acelaşi lucru s-a făcut chiar şi cu visele. Studiul care a realizat asta e cu totul uluitor:

> Imaginile vizuale din timpul somnului sunt de mult timp un subiect pe marginea căruia se tot speculează, dar natura lor privată împiedică analizele obiective. Prezentăm aici o abordare a decodificării neuronale în care modele de învăţare automată prezic conţinutul imaginilor vizuale din etapa de instalare a somnului, dată fiind activitatea cerebrală măsurată, prin descoperirea unor legături între tipare de RMNf la oameni şi rapoarte verbale cu ajutorul unor baze de date lexicale şi cu imagini. Modelele de decodificare formate în baza activităţii cerebrale induse de stimuli în ariile cortexului vizual au realizat clasificare, detectare şi identificare de conţinut precise. Concluziile noastre demonstrează că experienţa vizuală specifică din timpul somnului este reprezentată de tipare de activitate cerebrală distribuite de perceperea stimulului, oferind un

mijloc de dezvăluire a conţinutului subiectiv al visului prin utilizarea unei măsurări neuronale obiective[15].

De aici până la decodificarea gândurilor, operaţionalizate ca vorbire interioară sau ca imagini mentale, mai este un pas mic – foarte mic. Christian Herff şi colegii săi – care au arătat că „vorbirea rostită continuu poate fi decodificată în cuvintele exprimate [ca text scris] din trasee de electrocorticografie intracraniană (ECoG)” – spun aproape cu dezinvoltură că acesta va fi pasul următor în programul lor de cercetare: „Sistemul *Brain-to-Text* descris în această lucrare reprezintă un pas important spre comunicarea om-maşină bazată pe vorbirea imaginată”[16].

Asemenea evoluţii au implicaţii complexe. Să analizăm doar una, aparentă: dacă, în modul prevăzut în experimentul mental al lui Chalmers, este posibil să codificăm conţinutul proceselor corticale[17] într-un mediu artificial, apoi să le decodificăm înapoi în procesele fiziologice care le-au generat, atunci, în principiu, ar trebui să fie posibil să transferăm orice arie de procesare corticală într-un dispozitiv artificial şi să-l arhivăm, ca să-l utilizăm în viitor. Asta ar face posibil ca, în principiu, conţinutul cortexului unui individ să supravieţuiască corpului muritor al acestuia. Iar dacă asta se poate în cazul unui şoarece, de pildă, atunci ar trebui să se poată şi în cazul unei fiinţe umane – întrucât arhitectura de bază este aceeaşi.

Când am aflat de această posibilitate, iniţial nu m-a

entuziasmat[18]. Chiar dacă ar fi posibil să stochezi artificial întregul conţinut al memoriei pe termen lung a unei persoane, asta nu înseamnă că avem o copie a minţii ei. Astfel, am realiza doar puţin mai mult decât putem deja atunci când ţinem în viaţă în mod artificial corpul cuiva, de exemplu, într-o stare vegetativă sau în comă. Dacă aş face acest lucru pentru tine, nu te-aş menţine *pe tine* în viaţă, ci doar *cevaul* reprezentat de memorie. Plecând de la această analogie, dacă (sau poate ar trebui să spun când) ar deveni posibil să stocăm artificial conţinutul memoriei declarative pe termen lung – totalitatea oricărei cantităţi de parametri funcţionali individualizaţi ai cortexului cerebral –, tot nu am stoca decât un „ceva”, şi nu un „eu”. Raţionamentul meu s-a bazat pe opinia, fundamentată
ştiinţific, că parametrii memoriei corticale nu sunt intrinsec conştienţi. Nu ar fi *nimeni acasă* în aceste sisteme ale memoriei pe termen lung – nicio prezenţă subiectivă care să pipăie urmele cu ajutorul sentimentelor. Acesta este motivul pentru care am respins întregul demers privind inteligenţa artificială: „Dacă nu e posibil să proiectăm un calculator ce are *sentimente* [...], probabil nu va fi niciodată posibil să proiectăm un calculator ce are o minte [...]. Aşadar, problema minţii probabil că nu e o problemă a *inteligenţei”*[19].

Poate îţi dai seama în ce direcţie mă îndrept.

Până în 2002, când am scris aceste cuvinte, ştiam că conştiinţa este fundamental afectivă. (Uitându-mă peste lucrările mele, se pare că am ajuns treptat să văd aşa

lucrurile cam de prin 1996—.) Totuşi, pe atunci consideram că afectul nu are nicio legătură cu prelucrarea informaţiilor. Puneam semnul egalităţii între *cogniţie* şi prelucrarea informaţiilor şi consideram afectul ceva mai degrabă intrinsec biologic. Acum, privind în urmă – deşi încă mi se pare că aceste distincţii au noimă –, nu mai înţeleg de ce consideram că neuromodularea este mai „biologică” decât neurotransmisia. Mai mult, nu mai înţeleg de ce presupuneam că afectul nu este o formă de prelucrare a informaţiilor.

Poate că nu aveam cunoştinţe suficiente despre cibernetică. Mai exact, nu înţelegeam că *entropia* din mecanica statistică este echivalentă cu *informaţia medie* (vezi capitolul 7) – care înseamnă că tiparele de activitate cerebrală înregistrate la nivelul cortexului de cercetătorii menţionaţi mai sus sunt, în ultimă instanţă, tipare ale unor *deviaţii homeostatice* de la stările de repaus neuronal. Cu alte cuvinte, nu înţelegeam că şi activitatea corticală este entropică – că are aceasta în comun cu afectul – şi că travaliul cognitiv (adică travaliul predictiv) este *unti*-entropic şi, prin urmare, face parte din aceeaşi economie. Am început să înţeleg aceste lucruri abia în 2017, în timp ce îmi pregăteam remarcile de final la acel congres decisiv din Londra.

Acum, deşi până şi mie – noncognitivist declarat, „biolog al minţii” – mi se pare ciudat, sunt obligat să adopt o perspectivă diferită asupra inteligenţei artificiale. Astăzi, aş spune chiar că, dacă nu va fi posibil să facem o maşină conştientă, atunci nu vom

putea rezolva problema dificilă[21]. Dacă forma specială de prelucrare a informaţiilor pe care o propun aici este într- adevăr mecanismul cauzal al conştiinţei, atunci *trebuie* să fie posibil să producem în mod artificial o minte conştientă cu ajutorul ei. Acest rezultat final este singurul mod acceptabil din punct de vedere ştiinţific în care poate fi demonstrată această idee. Chiar dacă încă nu există (ca pe vremea lui Freud) tehnologia cu ajutorul căreia se poate testa o ipoteză, ipoteza tot trebuie formulată astfel încât să fie testabilă în principiu.

Şi, dacă vorbim de ipotezele pe care le-am conturat în această carte, tehnologia necesară *există.*

Mecanismul biologic al conştiinţei pe care l-am formulat în capitolul 6 este cea mai bună propunere a mea ca neuropsiholog. Se foloseşte într-un mod plauzibil de toate datele neurologice şi psihologice existente. S-ar putea adăuga multe alte elemente utile prin studierea mai atentă a mijloacelor fiziologice prin care PAG procesează numeroasele inputuri convergente pe care le primeşte şi a modulului în care se iau deciziile cu privire la „ce e de făcut în continuare” când PAG interacţionează cu coliculii superiori. Dar asemenea studii tot s-ar încadra în categoria pe care Chalmers o numeşte problema „simplă”.

În capitolul 9, cu ajutorul considerabil al lui Karl Friston, am redus funcţiile psihofiziologice la mecanica statistică formală. Ecuaţiile rezultate sunt nişte tuşe groase ale *mecanismelor cauzale* fundamentale ale conştiinţei în manifestările sale psihologice şi fiziologice

deopotrivă. Ingineria inversă a acestor abstracţiuni ne va spune dacă ele pot să rezolve problema dificilă. După cum arată lucrurile în prezent, există prea multe goluri între tuşele groase pe care le schiţez şi punerea în practică a mecanismelor descrise de ele; iar aceste etape lipsă vor deveni complet vizibile numai pe măsură ce vom încerca să le concretizăm.

Abia aştept să încep acest demers împreună cu echipa lărgită de fizicieni, informaticieni, specialişti în bioinginerie medicală şi în neuroştiinţe pe care am alcătuit-o. Până iese această carte din tipar, sper că vom putea raporta nişte progrese. Deocamdată, tot ce pot să spun este cum *preconizez* că vom proceda. Aşadar, tot ce urmează poate fi completat prin contribuţiile şi recomandările colegilor mei specialişti, pe măsură ce trecem de la o etapă la alta în proiectul nostru.

Mi se pare că punctul nostru de plecare va fi diferit de cel al majorităţii studiilor tradiţionale asupra „inteligenţei artificiale”. În primul rând, nu încercăm să construim prin inginerie inteligenţă, ci *conştiinţă*[22]. Sper că de-acum e clar că eu nu concep conştiinţa drept deosebit de inteligentă – cel puţin nu în forma sa elementară. Şi consider că e o strategie bună să încercăm să construim conştiinţa în forma sa elementară, nu doar din motive practice, ci şi etice, despre care voi vorbi imediat. În al doilea rând, nu vom încerca să construim un dispozitiv care face ceva practic – joacă şah, recunoaşte vocea sau altele asemenea –, iar când îi vom realiza criteriul funcţional va însemna că ne-am atins obiectivul. Mai degrabă, vom

construi un sistem care se autovalidează şi al cărui unic scop *obiectiv* este să rămână mijlocul de a atinge acest scop. Cu alte cuvinte, vom încerca să facem o fiinţă care nu are alt scop şi alt obiectiv decât să continue să fie[23].

Aşadar, vom începe cu un sistem similar celui creat de Friston: un sistem inconştient cu autoorganizare, echipat cu o zonă Markov (deci cu stări senzoriale, active şi interne) care creează automat un model al lumii în baza unor mostre senzoriale, minimizând efectele entropiei asupra integrităţii sale funcţionale prin îmbunătăţirea modelului său generativ. Adică, respectând legea lui Friston, îşi va măsura energia liberă preconizată şi va acţiona în consecinţă. Asta îl va transforma într-o maşinărie de predicţie. Apoi va păstra un model generativ ierarhic tot mai complex al organizării sale în relaţie cu stările externe predominante, chiar dacă, la fel ca Eve Periaqueduct, care astupa fisuri, nu va avea nicio altă sarcină explicită în afară de astuparea fisurilor din sistemul său.

Acest sistem va părea a fi însufleţit, dar nu va fi viu. Deşi conştiinţa a evoluat în organisme vii, obiectivul experimentului este să arate că ea poate fi produsă şi *artificial,* printr-o inginerie inversă a organizării sale funcţionale. Am toate motivele să cred că putem crea un asemenea sistem (cu autoorganizare), deoarece aşa ceva s-a mai făcut. Deşi într-o primă etapă acest sistem nu va fi conştient, el va avea deja valori subiective: stările sale valoroase (din propriul punct de vedere) vor fi cele care îi minimizează energia liberă şi el va face tot ce este necesar ca să şi le menţină cât mai mult. Protointenţionalitatea astfel implicată va fi relativ

simplu de verificat, întrucât poate fi inferată cu uşurinţă din comportamentul observabil al sistemului.

Ca să trecem în a doua etapă, în care cred că vor apărea mai întâi *precursorii* afectului, trebuie să ne facem sistemul mai complex. Mai exact, trebuie să-i dăm mai multe nevoi. Asta se poate face (şi aşa vom proceda iniţial) prin simulări pe calculator, dar o abordare mai realistă impune ca, în cele din urmă, *să întrupăm* sistemul din punct de vedere fizic, sub forma unui robot, şi să facem astfel încât capacitatea sa de a minimiza în mod proactiv energia liberă Friston să depindă de o sursă externă de energie liberă Gibbs[24]. Dacă am face acest lucru, sistemul ar trebui *să reprezinte* fondul de energie liberă Gibbs în termeni de energie liberă Friston, apoi să o fixeze prin lucru mecanic util de prelucrare a informaţiilor. Asta ar face menţinerea fondului său extern de energie o responsabilitate esenţială a sistemului cu autoorganizare. (Desigur, lucrul mecanic de prelucrare a informaţiilor desfăşurat de fiecare calculator depinde de electricitate, dar, în mod normal, aceasta este furnizată gratuit de un agent extern.) Astfel, hardware-ul calculatorului ar deveni corpul sistemului, chiar înainte de a-l pune într-un robot, iar valorile sistemului cu autoorganizare lar obliga să acţioneze – în cadrul propriului model al lumii – în moduri care îi permit cel mai bine corpului său să absoarbă energia externă necesară.

Iniţial, în simulare, am putea să proiectăm diverse artificii motoare care să permită fizic sistemului să acceseze sursa – sau, mai bine, sursele (la plural) – de

energie externă ca să-şi reîncarce bateria internă. De asemenea, am putea să transferăm sursele activate în locuri diferite din mediul calculatorului (folosind o rutină rezonabil de complexă ce i-ar fi necesară sistemului ca să înveţe). Apoi, asta ne-ar impune să echipăm sistemul cu un aparat perceptiv. În acelaşi spirit, am putea să-l încărcăm cu un mecanism de termoreglare, ca să nu se supraîncălzească dacă lucrează prea intens. Întrucât asta ar ameninţa integritatea fizică a corpului de care depinde sistemul cu autoorganizare, ar trebui să reprezinte şi să regleze şi acest homeostat. Aici am putea să încorporăm şi un parametru de epuizare, în termenii căruia eficienţa prelucrării informaţiilor de către sistem devine compromisă (treptat, progresiv) când este activ prea mult timp, fără repaus, iar acest parametru ar putea să interacţioneze în moduri complexe cu alte cereri pe care i le impunem. Am putea să adăugăm un parametru de durere, declanşat de deteriorarea (de exemplu, uzura) unor aspecte specifice ale integrităţii fizice a computerului, cum ar fi punctele de articulaţie şi suprafeţele roţilor şi braţelor. Asta ne permite să ne gândim şi la un parametru de anxietate, legat nu numai de pericolul perceput al unei dureri posibile, ci şi de riscul existenţial *iminent* referitor la alţi parametri. Ca să creăm emoţii mai realiste, va fi necesar şi să punem sistemul să concureze cu alţi agenţi (similari lui) pentru resurse energetice, să se confrunte cu ameninţări din partea lor, să aibă nevoi de ataşament faţă de ei şi să formeze alianţe cu ei. Şi aşa mai

departe. Acest lucru va concretiza problema emoţională importantă a nevoilor *contradictorii.*

Multe versiuni provizorii ale sistemului nu vor reuşi să satisfacă cererile împovărătoare imaginate mai sus şi, în consecinţă, vor sucomba. Prin urmare, va trebui să monitorizăm ce funcţionează cel mai bine şi să prestabilim codurile predictive de succes şi preciziile lor aferente pentru fiecare generaţie nouă de fenotip al sistemului. (Aceasta este „selecţia naturală” artificială, impusă de faptul că sistemele artificiale imaginate nu se vor autogenera şi nu se vor reproduce fizic[25].)

În sfârşit, ajungem în a treia etapă. Acum, numeroasele nevoi ale sistemului trebuie să fie prioritizate în mod flexibil printr-o optimizare a preciziei în funcţie de *context.* De exemplu, volumul energiei disponibile în exterior ar putea varia sub aspectul pragurilor termoreglatoare, de epuizare şi de anxietate, astfel încât evaluările optime ale preciziei dintr-un ciclu circadian sau cvasisezonier ar fi diferite de cele din alt ciclu. De asemenea, sistemului i s-ar putea permite să se adapteze la o configuraţie complexă de riscuri de mediu şi de oportunităţi, apoi parametrii ar putea fi modificaţi, creându-se astfel medii noi (neprevăzute) şi necesitatea unei planificări pe termen mai lung.

Fără îndoială, de aici va rezulta sucombarea multora dintre versiunile noastre de sistem cu autoorganizare. Aşadar, din aceleaşi motive ca mai înainte, ceea ce funcţionează cel mai bine va trebui propagat în mod artificial la nivelul mai multor generaţii succesive ale fenotipului. Nu e nevoie să fim exagerat de preocupaţi de acest lucru, deoarece sistemelor care nu

supravieţuiesc le lipseşte probabil ingredientul esenţial pe care încercăm să-l construim aici – capacitatea de conştientizare.

Conştiinţa (în sensul de incertitudine *simţită de sistem)* ar trebui să apară abia în această a treia etapă a experimentului. Sistemele cu autoorganizare impregnate cu valori de autoprezervare (cum şi sunt prin definiţie) conţin deja ingredientele brute pentru funcţia pe care o numim afect; adică deja consemnează „bunătate” şi „răutate” subiective de un tip valabil numai *în cazul lor* şi *pentru* ele şi, prin urmare, nu pot fi simţite decât *de* ele. Conform ipotezei mele, această proprietate inerent subiectivă este ceea ce numim *valenţă* hedonică. Exprimată în termeni formali, energia liberă în creştere este o criză existenţială pentru orice sistem cu autoorganizare. De aceea, asemenea sisteme trebuie să dezvolte modele interne ale lor înseşi în relaţie cu lumile lor, care le dau capacitatea de acţiune în scopul autoprezervării. Aşa ceva se califică drept comportament protointenţional. Dat fiind că complexitatea creşte, sistemul trebuie tot mai mult să-şi prioritizeze intenţiile în mod flexibil şi în funcţie de context, apoi să ţină „în minte” (într-o memorie-tampon pe termen scurt) o asemenea intenţie care să-i ghideze opţiunile în desfăşurare din medii nesigure. Va trebui şi să planifice în avans şi „să-şi amintească viitorul” pe intervale temporale mai lungi.

În acest moment, au şanse să supravieţuiască numai sistemele care pot să compartimenteze diferitele valori de eroare ce contribuie la calcularea energiei libere şi care îşi pot modula în mod flexibil cântăririle aferente

ale preciziei. Aceste valori trebuie tratate ca variabile categoriale, adică trebuie diferenţiate unele de altele pe criterii *calitative,* şi nu cantitative, iar una dintre aceste calităţi trebuie să fie prioritizată în orice moment, apoi implementată şi evaluată în termenii nivelurilor fluctuante de *încredere* ale sistemului. Predicţia mea este că, apoi, stările interne inerent subiective, având inerent valenţe, inerent existenţiale, inerent calitative, inerent intenţionale ale sistemului vor deveni ceea ce numim „sentiment”.

Nu văd niciun motiv pentru care aş ajunge la altă concluzie decât că un sistem suficient de complex, bine reglat, care se autovalidează şi e capabil să ajungă în această stare o poate *simţi.* Ceea ce ar simţi nu ar fi *totuna* cu ceea ce simt oamenii sau mamiferele sau animalele în general, însă ar putea fi sentimente. Acum, întrebarea este: cum vom şti dacă sau când sinele artificial se simte pe sine că este în această stare internă? Cum putem dovedi acest lucru?

Dacă vrei să-ţi faci o idee în linii mari despre actuala perspectivă generală asupra a ceva, Wikipedia este, evident, locul în care o poţi găsi. Iată ce spune despre conştiinţa artificială:

> Qualia sau conştiinţa fenomenologică reprezintă un fenomen în mod inerent la persoana întâi. Deşi diferite sisteme pot să manifeste diverse semne de comportament corelat cu conştiinţa funcţională, nu se poate concepe un mod în care teste la persoana a treia să aibă acces la caracteristici fenomenologice la

persoana întâi. Din această cauză şi din cauză că nu există o definiţie empirică a conştiinţei, un test al prezenţei conştiinţei în conştiinţa artificială ar putea fi imposibil[26].

Aceasta este o mare provocare. De fapt, este tot problema altor minţi. După cum am spus mai sus, nu încercăm să construim un dispozitiv cu inteligenţă artificială care face ceva practic, iar când îi vom realiza criteriul funcţional va însemna că ne-am atins obiectivul. Sistemul pe care trebuie să-l construim nu are alt scop obiectiv decât să supravieţuiască şi, mai exact, să supravieţuiască în medii neprevăzute. Dar, dacă îşi îndeplineşte criteriul obiectiv, cum vom şti dacă foloseşte *simţirea* în acest scop? La asta mă aştept – previziunea mea este că va folosi sentimente –, dar sentimentele sunt inerent subiective, aşa că, după cum ne aminteşte Wikipedia, „nu se poate concepe un mod” în care să le demonstrăm obiectiv prezenţa.

Din nou, regulile ştiinţei ne vin în ajutor. Ştiinţa nu necesită dovezi absolute. Această libertate nu e valabilă numai în cazul mecanismului conştiinţei, ci şi în cazul a orice. Regulile ştiinţei cer numai să propui o ipoteză rezonabilă (care ţine cont de toate dovezile disponibile), apoi să o testezi într-un mod ce asigură o posibilitate realistă să fie falsificată. Aşadar, în cazul nostru, când vom şti că e falsificată ipoteza că sistemul nostru ce se autovalidează este conştient? După cum am discutat în capitolul anterior, se poate concepe că există „zombi afectivi” care *se comportă* ca şi cum ar fi conştienţi, dar nu sunt conştienţi. Atunci, cum putem şti vreodată dacă

ei *chiar sunt* conştienţi?

Răspunsul tradiţional la această întrebare este testul Turing, un test foarte simplu care îţi cere să interacţionezi cu sistemul prin intermediul unui ecran de calculator (adică fără să-l *vezi)* şi să stabileşti numai în funcţie de afirmaţiile sale verbale dacă are minte sau nu. Dacă tu crezi că interacţionezi cu o minte reală (din punct de vedere tehnic, dacă 30% dintre evaluatori cred că da, după cinci minute de interacţiune), atunci maşina a trecut testul.

Iată un exemplu pe care l-am publicat împreună cu Oliver Turnbull în 200227. El a interacţionat – online – cu cineva sau ceva pe nume ELIZA, care susţinea că este psihoterapeut. Sarcina lui era să stabilească dacă ea este o persoană reală sau nu:

ELIZA: Bună! Spune-mi, te rog, care e problema ta.

Oliver: De ce nu pot să fiu fericit?

ELIZA: Crezi că ar trebui să poţi să fii fericit?

Oliver: Cred că merit să fiu mai fericit decât sunt acum.

ELIZA: Chiar crezi asta?

Oliver: Da, nu e nevoie ca viaţa să fie atât de dură.

ELIZA: Pari că ai o atitudine pozitivă.

Oliver: Presupun că da, cel puţin o parte a timpului.

ELIZA: Şi de asta ai venit la mine?

Oliver: Am venit pentru că am nevoie de ajutor.

ELIZA: Acesta e adevăratul motiv?

Oliver: De fapt, am venit pentru că am auzit că eşti un calculator, nu un terapeut în carne şi oase.

ELIZA: Te îngrijorează calculatoarele?

Oliver: Pot să-ţi pun o întrebare?

ELIZA: Dacă poţi să-mi pui o întrebare depinde mai mult de tine decât de mine.

Oliver: Întrebarea este: cum relaţionează mintea şi creierul?

ELIZA: Nu sunt sigură că înţeleg tot ce vrei să spui.

Oliver: Asta e problema dificilă – cum apare viaţa psihică din evenimente fiziologice?

ELIZA: Continuă, te rog…

Sper că ai ghicit că ELIZA nu este o persoană reală, ci un chatbot – un program de calculator – şi încă unul foarte simplu. Dar un număr surprinzător de mare dintre oamenii care interacţionează cu ea cred că chiar e o terapeută şi mulţi dintre ei o consideră utilă[28]. În mod clar, ceva nu e în regulă cu testul Turing. După cum am afirmat eu şi Oliver Turnbull, principalul său neajuns constă în faptul că este un test *behaviorist* de inteligenţă, şi nu un test de *conştiinţă.* E adevărat că comportamentul inteligent al unei maşini poate fi uneori imposibil de deosebit de cel al unei persoane, iar când se întâmplă aşa, maşina trebuie tratată ca fiindu-ne egală (sau chiar superioară) în ce priveşte inteligenţa[29]. Adică atât oamenii, cât şi calculatoarele manifestă „inteligenţă”. Dar din problema zombilor filosofici vine alta – mai precis, *nu* e vorba de comportament sau de inteligenţă, ci de faptul dificil că, dacă ne referim la conştiinţă, aparenţele pot fi înşelătoare.

De-a lungul anilor s-au mai propus teste formale – unele dintre ele vizând în mod specific testarea conştiinţei –, dar toate sunt pur şi simplu la fel de nepotrivite pentru obiectivele noastre ca testul Turing,

în principal din cauză că pleacă de la asumpţia că încerci să depistezi conştiinţă cognitivă, nu afectivă[30]. Noi încercăm să construim ceva mult mai simplu: o minte a cărei procesare cognitivă nu este mai complexă decât e necesar ca să-şi simtă rezervele de energie împuţinându-se ori creşterea excesivă a temperaturii.

Partea bună a testului Turing este că ocoleşte prejudecăţile[31]. Există pericolul ca noi, oamenii, să nu presupunem a priori că ceva ce arată ca o „simplă maşină” nu poate fi conştient. Asta se poate transforma într-o profeţie care se autoîmplineşte. Omenirea are un lung istoric de asemenea manifestări ale intoleranţei, care încă nu s-a încheiat. Aici nu mă refer în special la prejudecăţile pe motive de rasă, gen sau orientare sexuală, care sunt şi ele suficient de rele, ci la asumpţia că copiii născuţi fără cortex trebuie să fie inconştienţi, ca şi animalele nonumane. Dacă mulţi oameni, chiar reputaţi specialişti în neuroştiinţe, tind să creadă că şobolanii – care sunt tot mamifere, ca şi noi, echipaţi cu, practic, aceeaşi anatomie a mezencefalului ca noi şi cu un cortex, ale cărui acţiuni susţin ipoteza că sunt conştienţi – nu au totuşi conştiinţă, atunci ce speranţă putem avea să accepte ideea că sinele nostru artificial este capabil să simtă, indiferent câte dovezi aducem?

Va trebui să vedem ce se întâmplă. În ce mă priveşte, tot ce pot să fac este să-mi expun explicit predicţiile şi să spun cum intenţionez să le testez. Principala mea predicţie este că sistemul din a două etapă descris mai sus nu va fi capabil să supravieţuiască în medii noi, spre deosebire de sistemul din a treia etapă (sau unele dintre

versiunile lui). Acesta este criteriul meu operaţional pentru activitatea *voluntară* (vezi p. 109 pentru definiţia termenului „voluntar”). Mai mult, prezic că cele două rezultate finale vor coincide cu vreun aspect esenţial al funcţionării unui mecanism de *prioritizare a nevoilor* (adică optimizarea preciziei), pe care îl va avea numai sistemul din a treia etapă. Care anume va fi trăsătura esenţială în cadrul acestui mecanism amplu se poate determina numai prin încercare şi eroare. Pe scurt, trebuie să identificăm „corelativul neuronal artificial” al conştiinţei al sistemului nostru – mecanismul său de selecţie a afectului şi cel prin care păstrează în minte un afect prioritizat şi îl foloseşte ca să califice incertitudinea într-o secvenţă de acţiune în desfăşurare. Identificarea unei asemenea trăsături ne va permite să o manipulăm, în mare parte cum am procedat cu componentele creierului de vertebrate despre care am concluzionat că sunt responsabile pentru conştiinţa *noastră* (ce rămâne unica formă de conştiinţă pe care o putem verifica direct, empiric, adică în cazul fiecăruia dintre noi, din cauza problemei altor minţi).

De exemplu, putem prezice cu încredere că *deteriorarea* corelativului neuronal al conştiinţei în sistemul nostru îi va anula conştiinţa cam tot aşa cum se întâmplă în cazul leziunilor de la nivelul nucleilor parabrahiali la vertebrate, cum suntem şi noi, sau mai degrabă că deteriorarea îi va afecta comportamentul voluntar aşa cum se întâmplă în cazul leziunilor la nivelul PAG la vertebrate. De asemenea, putem prezice că *stimularea* (adică amplificarea) acestei componente esenţiale a sistemului va facilita activitatea voliţională.

Desigur, ne putem aştepta ca activitatea internă înregistrată din partea sistemului să prezică nu numai evenimente externe, ci şi comportamentele voluntare aferente orientate spre obiective şi diferite aspecte ale acestei activităţi înregistrate să corespundă unor aspecte diferite ale comportamentelor observabile.

Cel mai mult sper că, după ce vom fi identificat corelativul neuronal al conştiinţei în sistemul nostru artificial, această componentă se va dovedi a fi suficient de distinctă în comparaţie cu alte componente ale arhitecturii funcţionale a sistemului – mai exact, cele responsabile cu implementarea comportamentelor adaptive generate în mod normal de sentimente – încât să putem manipula sentimentele presupuse *decuplate* de la consecinţele lor adaptive. Să ne amintim ce am spus în capitolul 5 despre deosebirea dintre motivaţiile psihologice subiective şi principiile obiective ale designului biologic. De exemplu, comportamentul sexual este motivat de regulă de plăcerea pe care o produce, şi nu de imperativele reproducerii, care, pe parcursul evoluţiei, au ataşat „recompensa” biologică actelor de procreaţie. Aici am în vedere ceva analog cu ceea ce se observă la persoanele cu dependenţe, care sunt motivate să facă efort ca să obţină sentimentele dorite, deşi sentimentele în sine nu dau niciun avantaj adaptiv sistemului în ceea ce priveşte principiile fundamentale ale designului său[32]. Se poate observa ceva similar şi la animalele (peştii zebră, de exemplu) care manifestă comportamente ce au fost condiţionate de preferinţe pentru anumite locuri în care au primit opiacee, cocaină, amfetamine şi nicotină – adică *recompense*

hedonice care dau prea puţine avantaje adaptive sau chiar deloc şi care pot de fapt să le dăuneze[33]. Mi se pare că, dacă putem demonstra ceva echivalent pentru sistemul nostru, acest ceva va fi o dovadă hotărâtoare că există sentiment subiectiv – dovadă care ar putea fi apoi supusă validării încrucişate cu ajutorul manipulărilor cauzale (cum ar fi deteriorarea şi stimularea artificiale) şi al tehnicilor de înregistrare prevăzute mai sus.

Bineînţeles, în tot acest context tot se va pune problema altor minţi. Dar acelaşi lucru e valabil şi în cazul meu şi al tău.

Eu nu pot şti niciodată sigur dacă tu eşti conştient. În cele din urmă, totul se reduce la concluzii convergente şi la ponderea dovezilor. Poate că nu va exista niciodată un consens cu privire la acest subiect, la fel cum astăzi nu există un consens dacă copiii cu hidroanencefalie şi animalele nonumane sunt subiecte ale experienţei sau nu. Oamenii care nu acceptă că aceşti semeni ai noştri sunt conştienţi probabil nu vor accepta niciodată că o „fiinţă” artificială simte ceva, indiferent câte dovezi se aduc în acest sens. Pentru noi, ceilalţi, ar trebui să existe întotdeauna loc de îndoială. Cât despre mine, dacă văd că sunt semnificativ *indecis* cu privire la întrebarea dacă sinele nostru artificial este conştient, acesta ar fi un rezultat remarcabil.

Ar trebui să facem asta? Când echipa noastră a început să se gândească la un proiect de cercetare de tipul pe care tocmai lam descris, în scurt timp ne-au copleşit considerentele etice. Care este rostul

construirii unei asemenea maşinării? Cine ar putea beneficia de pe urma ei şi în ce fel, cu ce costuri şi în contul cui? Pe scurt, care sunt riscurile?

Multe proiecte şi aplicaţii actuale de inteligenţă artificială se fac din considerente comerciale. Oare să acceptăm finanţare pentru cercetare din partea cuiva care urmăreşte profitul financiar prin proiectul nostru? Pe ce criterii un dispozitiv conştient în mod artificial ar putea deveni profitabil pentru cineva?[34] E de înţeles că unele persoane cu motivaţii comerciale (în sens juridic, adică inclusiv corporaţiile) ar putea dori să înlocuiască forţa de muncă umană cu unităţi de producţie artificiale, în care acestea din urmă pot fi sau mai eficiente decât noi – inclusiv mai capabile din punct de vedere intelectual –, sau mai „dispuse” decât noi să îndeplinească sarcini monotone şi continue. Până şi acest motiv este discutabil dintr-o perspectivă etică, în măsura în care generează îngrijorări cu privire la şansele de angajare reduse ale oamenilor, dar cel puţin nu se poate pune problema *exploatării* unor maşini nonconştiente.

Dar asta nu se poate spune şi în cazul proiectului nostru. În măsura în care conştiinţa artificială poate fi utilizată pentru câştiguri financiare, riscăm să facilităm o formă nouă de sclavie. Iar asta ar fi o dovadă de lipsă de empatie crasă, cum a fost întotdeauna. Aşadar, nu pot să-mi imaginez nicio justificare etică pentru dezvoltarea unor roboţi capabili să simtă printr-un program de cercetare cu finanţare comercială – şi nici chiar pentru dezvoltarea lor indiferent până în ce etapă pentru asemenea scopuri practice, dacă asta creează

posibilitatea exploatării lor în acest fel.

Bineînţeles, preocuparea noastră pentru bunăstarea unor maşini posibil conştiente ar trebui să meargă dincolo de teama că ar putea fi exploatate din raţiuni pecuniare. Odată ce maşinile dobândesc conştiinţă (chiar şi cea mai rudimentară formă de senzaţie brută), apar cu necesitate întrebări mai generale referitoare la potenţialul lor de *a suferi.* Tradiţia dominantă în teoria eticii occidentale, „consecvenţialismul” (adică ideea că consecinţele comportamentului unei persoane sunt baza supremă pentru a stabili dacă este corect sau greşit), e profund preocupată de durere şi suferinţă. Prin urmare, prin crearea unor fiinţe artificiale care simt intrăm în jurisdicţia acestui tip de calcul etic.

Problema *drepturilor* este şi mai controversată. Oare, odată ce maşinile devin fiinţe capabile să simtă, problemele dezbătute în prezent sub titluri precum „drepturile omului”, „drepturile animalelor”, „drepturile copilului”, „dreptul la viaţă” şi aşa mai departe devin aplicabile şi în cazul lor? Oare maşinile conştiente trebuie să aibă drepturi la „viaţă” şi la libertate? De fapt, conceptul de „drepturi ale roboţilor” este deja consacrat, iar această chestiune este în analiza Institute for the Future din SUA şi a Ministerului pentru Comerţ şi Industrie din Marea Britanie[35].

Cu titlu de exemplu, voi menţiona doar două probleme valabile în cazul roboţilor conştienţi (spre deosebire de cei inteligenţi). Când încercăm să-i creăm – chiar şi numai pentru a afla dacă aşa ceva e posibil –, avem oare motive întemeiate să-i punem în mod intenţionat în situaţii cu potenţial stresant, ca să

demonstrăm că există reacţii de evitare? (Această întrebare apare adesea în cazul experimentelor pe animale.) Şi, presupunând că putem crea maşini capabile să simtă, pe ce temeiuri etice s-ar justifica să le oprim? Şi nu e deloc greu să invocăm şi mai multe exemple.

Pe lângă asemenea întrebări etice şi morale, trebuie luate în considerare şi chestiuni practice, unele dintre ele foarte importante, chiar existenţiale. De exemplu, întrucât calculatoarele sunt deja mai inteligente decât noi sub anumite aspecte limitate, oare nu e posibil ca cele care sunt *şi* foarte inteligente, *şi* conştiente să dezvolte motivaţii care ar putea să nu fie în interesul omenirii? Această posibilitate e de multă vreme o preocupare a scriitorilor şi futurologilor cu imaginaţie bogată, dar aş vrea să atrag atenţia în mod deosebit asupra faptului că conştiinţa, aşa cum am ajuns să o înţelegem în cartea de faţă, este, spre deosebire de inteligenţă, profund legată de credinţa că e „bine” să supravieţuieşti şi să te reproduci. Prin urmare, o maşină inteligentă al cărei comportament se înrădăcinează în acest sistem de valori prezintă un pericol special nu numai pentru oameni, ci poate şi pentru toate celelalte forme de viaţă. Desigur, acest pericol apare pentru orice formă de viaţă considerată a fi o ameninţare potenţială pentru aceste maşini care au un sine sau chiar un concurent important pentru resursele lor. Inteligenţa combinată cu motivaţia de autoprezervare este ceva diferit de simpla inteligenţă.

Nu voi enumera toate preocupările etice şi pericolele potenţiale ce apar odată cu posibilitatea conştiinţei

artificiale. Deja s-a scris mult pe acest subiect. Faptul că asemenea întrebări sunt tratate cu seriozitate de atâţia oameni având poziţii influente spune multe despre situaţia actuală a inteligenţei artificiale[36]. Chiar şi numai acest fapt te-ar putea încuraja să priveşti cu mai multă îngrijorare posibilitatea de a exista maşini capabile să simtă. Desigur, ţin seama mult mai mult de asemenea chestiuni decât până acum. Nici măcar în 2017, nu consideram realizabilă conştiinţa robotică, nu numai în timpul vieţii mele, ci în principiu. Între timp, m-am răzgândit.

Deci, având în vedere toate aceste preocupări etice, de ce consider că e necesar să încerc să demonstrez că conştiinţa se poate produce artificial? Pur şi simplu, pentru că asta mi se pare a fi singura cale prin care poate fi falsificată ipoteza avansată în cartea de faţă. Dacă şi până când nu construim conştiinţă, nu putem fi încrezători că am rezolvat problema referitoare la motivele pentru care *şi felul în care* apare ea.

E acest motiv suficient ca să-mi asum riscul crucial pe care tocmai l-am descris? Răspunsul meu la această întrebare porneşte de la următoarea convingere: *dacă ceva se poate face, atunci va fi făcut.* Cu alte cuvinte, dacă, în principiu, conştiinţa poate fi construită prin inginerie, atunci asta se va întâmpla undeva, la un moment dat[37]. Această predicţie este valabilă indiferent dacă ipotezele avansate în cartea de faţă sunt corecte sau nu. Totuşi, eu sunt răspunzător de ipotezele de aici şi de posibilitatea ca ele să fie corecte. Dacă sunt corecte sau măcar sunt pe drumul cel bun, atunci crearea conştiinţei artificiale este iminentă. Cu alte

cuvinte, asemenea ipoteze *vor fi* folosite în scurt timp pentru a construi conştiinţă.

Faptele individuale care au dus la concluziile prezentate în această carte – aproape toate – sunt publice de mai mulţi ani. Deşi e adevărat că mulţi specialişti în neuroştiinţe le interpretează altfel decât mine, este adevărat şi că alţii au ajuns la concluzii foarte similare. Chiar dacă fiecare dintre ei a evidenţiat aspecte diferite şi le-a nuanţat diferit, e corect să spun că Jaak Panksepp, Antonio Damasio şi Björn Merker, cel puţin, au ajuns la concluzia că conştiinţa (1) este generată în trunchiul cerebral superior, (2) este fundamental afectivă şi (3) este o formă extinsă a homeostazei. Aceste fapte combinate înseamnă că conştiinţa nu e atât de complicată pe cât am crezut până acum. Deci e rezonabil să ne aşteptăm să o putem construi. Singura completare importantă adusă de cartea de faţă la aceste concluzii este (4) principiul energiei libere. În esenţa lui, nici acesta nu e foarte complicat. De fapt, e tentant pentru că reduce aproape toate procesele psihice şi neurologice la un singur mecanism, făcându-le calculabile.

Şi principiul energiei libere este deja în domeniul public. Mai mult, eu şi Friston am publicat deja articole ştiinţifice în care combinăm acest principiu cu celelalte trei enumerate[38]. Până la urmă, ar fi neobişnuit să exprimi asemenea ipoteze într-o carte destinată publicului larg înainte să le supui evaluării unor colegi şi să le publici în reviste de specialitate relevante. Acelaşi lucru e valabil şi în cazul prelegerilor, când ţi se cere să-ţi aperi afirmaţiile în forumuri ştiinţifice şi

academice. Eu am prezentat ideile din această carte în faţa mai multor audienţe din lume, de diferite specializări[39].

Aşa că m-am dat de gol. Înseamnă asta că ar fi trebuit să mă abţin de la publicarea acestor articole şi de la prelegeri? Răspunsul este, fără echivoc, nu. Dacă nu aş fi făcut-o eu, ar fi făcut-o altcineva. Aceste idei se fac simţite. Le-a venit vremea. Nu încerc să mă apăr; e un fapt demonstrabil. Gândeşte-te, de exemplu, la articolul colegilor noştri cehi care au preconizat în 2017 că o soluţie la problema dificilă este iminentă şi că s-ar baza pe principiul energiei libere[40]. Robin Carhart-Harris (care a participat la unele dintre primele întâlniri despre neuropsihanaliză ţinute la Londra) a publicat în mod independent idei similare[41]. La fel şi Katerina Fotopoulou, specialistă în neuroştiinţe sociale, care a recunoscut încă din 2013 legătura dintre incertitudine (precizie inversă) şi conştiinţă[42]. Fără îndoială, dacă problema dificilă a conştiinţei urmează să fie rezolvată printr-o combinare între cele patru idei enumerate mai sus, atunci asta nu are cum să nu se întâmple în viitorul foarte apropiat, cu sau fără implicarea mea.

Pe această realizare se bazează abordarea pe care m-am hotărât s-o adopt în legătură cu chestiunile etice şi morale ce ne preocupă. Abordarea mea este opusul păstrării secretelor şi presupune acordarea de sprijin oricând e posibil, pentru aplicarea practică a ipotezelor mele – şi încă fără ezitare. Această abordare decurge logic din aşteptarea că, dacă ceva se poate face, atunci va fi făcut. De aceea, trebuie să mă situez înaintea

valului, astfel încât să pot preveni posibilele consecinţe nocive, ori de câte ori este posibil.

În esenţă, asta înseamnă că proiectul schiţat în acest capitol trebuie implementat *acum* şi fără finanţare comercială. În cazul în care echipa mea de cercetare reuşeşte să satisfacă criteriul prezentat mai sus – supravieţuirea sinelui artificial în medii neprevăzute – şi că obţinem dovezi rezonabile că este capabil să simtă (adică dacă nu există o *infirmare* a predicţiilor mele în acest sens), atunci consider că trebuie să trecem imediat la următoarele trei etape.

Cred că mai întâi trebuie să oprim maşina şi să-i scoatem bateria internă. Îmi dau seama că astfel se pune în discuţie una dintre preocupările etice enumerate mai sus, dar cred că e corect să începem cu asta. Să ne amintim că o maşină de tipul celei preconizate aici *nu va fi vie.* Nu văd de ce oprirea unei maşini conştiente nonvii ar presupune sucombarea ei. Tot timpul trebuie să fie posibil să o repornim şi, probabil, agentul conştient astfel reactivat va fi identic cu cel care a fost oprit (folosind analogia biologică a somnului şi trezirii). De observat că acest plan de a opri maşina respectă clauza referitoare la obligaţia de oprire inclusă în recomandările universale privind inteligenţa artificială (2018), ce reprezintă „declaraţia supremă de răspundere pentru un sistem cu inteligenţă artificială”[43].

În al doilea rând, trebuie să demarăm procesul de brevetare a componentei esenţiale a sistemului nostru din a treia etapă – corelativul neuronal al conştiinţei, indiferent ce se dovedeşte a fi – care ne-a permis să ne

atingem criteriile declarate. Nu se pot breveta ecuaţii, deci nu există riscul ca altcineva să o facă în timp ce noi încercăm să le aplicăm. Totuşi, întrucât ecuaţiile sunt deja în domeniul public, e imperativ să acţionăm repede, astfel încât să putem controla punerea lor concretă în aplicare până nu ne-o ia altcineva înainte. Dacă se ajunge la asta, e important ca brevetul să fie înregistrat pe numele unei organizaţii nonprofit corespunzătoare – precum OpenAI sau Future of Life Institute –, şi nu al unei persoane sau al unui grup de persoane. Cel puţin, acest lucru asigură luarea deciziilor în grup şi creşte şansele ca ele să fie în interesul binelui comun.

În al treilea şi ultimul rând, dacă ne atingem criteriile şi înregistrăm brevetul, proprietarii lui ar trebui să organizeze un simpozion în cadrul căruia specialişti, filosofi şi alte părţi interesate importante să fie invitate să-i analizeze implicaţiile şi să facă recomandări pentru paşii următori, inclusiv dacă, când şi în ce condiţii maşina capabilă să simtă trebuie repornită – şi, posibil, dezvoltată mai departe. Să sperăm că acest lucru va duce la elaborarea unui set de recomandări şi constrângeri mai ample asupra dezvoltării, exploatării şi proliferării viitoare a inteligenţei artificiale capabile să simtă în general.

Cred că toate acestea trebuie spuse, oricât de surprins sunt eu însumi că am făcut asemenea recomandări. Dar să nu ne facem iluzii cu privire la fragilitatea lor. Cunoaştem cu toţii precedentul energiei atomice şi al armelor nucleare. Singura opţiune este să facem ce putem, cât mai curând, pentru a recunoaşte

scara implicaţiilor ce rezultă din capacitatea noastră iminentă de a crea maşini conştiente.

Acum pare că avem la îndemână un criteriu obiectiv suficient de clar al conştiinţei. Sperăm că ne va influenţa comportamentul etic într-un sens mai general, dincolo de întrebările limitate ce rezultă din posibilitatea de a crea maşini conştiente. În general, se consideră că sentimentele sunt condiţii necesare şi suficiente pentru preocupările etice. Prin urmare, înţelegerea ştiinţifică a sentimentelor schiţată în această carte ne dă ocazia să analizăm mai profund suferinţa *animalelor.* Am menţionat nu o dată că progresele în neuroştiinţele afective de la sfârşitul secolului XX (adică înţelegerea că pentru a fi o fiinţă capabilă să simtă nu e nevoie decât de un triunghi de decizie în mezencefal, lucru pe care noi îl avem în comun cu toate vertebratele) au schimbat opiniile multor specialişti despre ce este acceptabil şi ce nu în studiile pe animale. Pare de la sine înţeles că acelaşi lucru ar trebui să se aplice şi atitudinii tuturor oamenilor faţă de bunăstarea animalelor în general. De exemplu, cum justificăm creşterea şi sacrificarea la scară industrială a altor fiinţe sensibile, ca să le mâncăm? Când discutăm despre asta, să nu uităm că conştiinţa apare gradual, astfel încât nu putem pune semnul egalităţii între presupusa conştiinţă a unei muşte sau a unui peşte şi cea a unui om. Totuşi, în aceeaşi logică, să nu uităm că oile, vacile şi porcii (care se regăsesc peste tot în meniurile occidentale) sunt *mamifere,* cum suntem şi noi. Asta înseamnă că pot trăi

aceleaşi emoţii de bază ca şi noi, precum FRICA, PANICA/MÂHNIREA şi GRIJA. Şi mamiferele au un cortex, adică sunt capabile – toate, într-o anumită măsură – „să-şi amintească viitorul” în mod conştient şi să se orienteze printre probabilităţile şi verosimilităţile acestuia în funcţie de ceea ce simt.

Suntem în secolul XXI şi, în lipsa unui obiectiv superior – dacă nu suntem decât conştiinţa noastră –, ce să facem dacă nu să minimizăm suferinţa? Acum, că ştim mai bine unde ar putea să existe suferinţă, ce mai putem face cu această informaţie? În mod clar, menţinerea şi protejarea conştiinţei biologice nu sunt legate doar de soarta speciei noastre.

Având în vedere tot ce te-am rugat să abandonezi pe parcursul cărţii de faţă – în ce priveşte excepţionalismul uman şi altele asemenea –, mi se pare potrivit să închei cu o scurtă reflecţie la ceea ce am putea câştiga în legătură cu imaginea noastră despre noi înşine din aceste introspecţii neplăcute.

Sentimentul este o moştenire preţioasă. Poartă în sine înţelepciunea veacurilor: o moştenire ce datează de eoni, de la începuturile vieţii. Când, în cele din urmă, homeostaza a dat naştere sentimentelor, aspectul esenţial al acestei capacităţi noi a fost că ne-a permis să ştim *cum ne descurcăm* pe o scară biologică de valori. Sentimentele antrenează predicţii înrădăcinate în experienţele acumulate în situaţii de importanţă biologică ale tuturor strămoşilor noştri. Sentimentele ne permit să facem ce este cel mai bine pentru noi, chiar şi atunci când nu ştim *de ce* facem aşa. Te-am mai rugat să-ţi imaginezi ce s-ar întâmpla dacă fiecare dintre noi

ar trebui să înveţe din nou ce alimente conţin o cantitate mare de energie şi dacă ar trebui să descoperim singuri ce se întâmplă când sari de pe o stâncă. Datorită sentimentelor spontane care ne împing spre dulce şi ne fac să evităm înălţimile, „pur şi simplu ştim” (la o primă aproximare) ce este de făcut şi când. De exemplu, ştim ce este de făcut când plâng bebeluşii, când atacă prădătorii sau când ne ies în cale obstacole frustrante. Această cunoaştere înnăscută – care ne este transmisă explicit *numai* sub formă de sentimente – ne permite să supravieţuim în lumile foarte imprevizibile în care trăim, unde pe lângă noi gonesc autovehicule, iar aerul se umple de dioxid de carbon.

Aşadar, pe măsură ce abandonăm iluzia familiară că conştiinţa circulă prin simţurile noastre, precum şi concepţia greşită conform căreia conştiinţa e sinonimă cu înţelegerea, să ne consolăm cu faptul că ea vine spontan din straturile noastre intime. Apare în noi chiar înainte să ne naştem. La sursa ei, suntem ghidaţi de un flux constant de sentimente ce provin dintr-un izvor de intuiţie, apărut nu se ştie de unde. Nu cunoaşte fiecare dintre noi, în mod individual, cauzele, dar le simţim. Sentimentele sunt o moştenire pe care ne-o lasă întreaga istorie a vieţii, ca să ne călească în faţa incertitudinilor viitoare.

Post-scriptum

La scurt timp după ce am scris o versiune completă a acestei cărţi, am fost invitat să prezint teza principală în cadrul congresului anual Ştiinţa Conştiinţei, care a avut loc la Interlaken, în 2019. Pentru asta a trebuit să condensez mai tot ce tocmai ai citit în formatul unei prelegeri plenare. Ar putea fi util să închei lunga noastră călătorie printr-un rezumat al celor 13 idei pe care le-am folosit pentru prelegerea respectivă.

(1) Marele fiziolog din secolul al XIX-lea Johannes Müller considera că organismele însufleţite „conţin un oarecare element nonfizic sau sunt guvernate de alte principii decât lucrurile neînsufleţite”. Studenţii lui (Helmholtz, Brücke, Du Bois-Reymond, Ludwig şi alţii) nu erau de acord, ci erau siguri că „în organism nu acţionează nicio altă forţă în afară de cele fizice şi chimice obişnuite”. La rândul său, elevul lor Sigmund Freud a încercat să întemeieze o ştiinţă naturală a minţii pe această bază, în care viaţa psihică poate fi redusă la „stări determinate cantitativ ale unor particule materiale detectabile”. A eşuat în acest proiect, lipsindu-i metodele, şi l-a abandonat în 1896.

(2) Un secol mai târziu (1994), biologul inovator Francis Crick a declarat că „tu, bucuriile şi tristeţile tale, amintirile şi ambiţiile tale, sentimentul identităţii tale personale şi al liberului-arbitru, nu esti de fapt decât comportamentul unui mare ansamblu de celule nervoase, cu moleculele lor aferente”. Crick ne-a îndemnat să încercăm din nou să descoperim corelativele neuronale ale conştiinţei, direcţie în care şi-a îndreptat şi el eforturile. Din păcate, a folosit însă constiinţa vizuală ca model.

(3) În replică, filosoful David Chalmers a afirmat că preocuparea lui Crick pentru corelativele neuronale ale constiinţei era o problemă „usoară” – una corelaţională, şi nu cauzală –, a cărei soluţie putea să explice unde, dar nu şi de ce Şi cum apare constiinţa. Pentru Chalmers, problema „dificilă” a constiinţei era: cum şi de ce activităţile neurofiziologice produc experienţa constiinţei? Pentru el (şi pentru predecesorul lui filosofic Thomas Nagel), problema se învârtea în jurul a ceva- cum-e al experienţei: „Un organism are stări mentale constiente dacă şi numai dacă există ceva cum e să *fii* acel organism – ceva cum e *pentru* acel organism”. Prin urmare, problema dificilă este: de ce şi cum apare calitatea subiectivă a experienţei din evenimente neurofiziologice obiective?

(4) E cam vag să întrebi cum anume niste lucruri obiective *produc* niste lucruri subiective – asemenea termeni riscă să facă problema dificilă mai dificilă decât e necesar. Obiectivitatea şi subiectivitatea sunt perspective observaţionale, nu cauze şi efecte. Evenimentele neurofiziologice nu pot să producă

evenimente psihologice tot aşa cum fulgerele nu pot să producă tunete. Ele sunt manifestări paralele ale unui singur proces de bază. Cauza fundamentală atât a fulgerului, cât şi a tunetului este electricitatea, ale cărei mecanisme legitime le explică pe amândouă. La fel, şi fenomenele fiziologice şi psihologice pot fi reduse la cauze unitare, dar nu unele la celelalte.

(5) De obicei, descriem cauzele fundamentale ale fenomenelor biologice în termeni „funcţionali”, iar mecanismele funcţionale se pot reduce, la rândul lor, la legi naturale. De exemplu: care este mecanismul vederii? Totuşi, Chalmers face observaţia corectă că mecanismul funcţional al vederii nu explică cum e să vezi. Asta pentru că vederea nu este o funcţie intrinsec conştientă. Nu e nevoie ca îndeplinirea funcţiilor vizuale (nici chiar a celor specific umane, precum cititul) să se simtă în vreun fel. Percepţia se întâmplă rapid, fără să conştientizezi ce este perceput, iar învăţarea fără să conştientizezi ce se învaţă. Prin urmare, Chalmers a pus, pe bună dreptate, întrebarea: „De ce îndeplinirea acestor funcţii este însoţită de experienţă? De ce toată această prelucrare a informaţiilor nu continuă «pe întuneric», fără să se simtă nimic în interior?”. Eşecul ştiinţei de a răspunde la această întrebare creează posibilitatea ca conştiinţa să nu facă parte din matricea cauzală obişnuită a universului.

(6)Întrebarea lui Chalmers poate fi pusă în legătură cu toate funcţiile cognitive, nu numai cu cele vizuale, dar nu şi cu funcţiile afective. Cum poţi să ai un

sentiment fără să-l simţi? Cum putem explica mecanismul funcţional al afectului fără să explicăm de ce şi cum ne face să experiem ceva? Până şi Freud era de acord în această privinţă: „Din esenţa unui sentiment face totuşi parte faptul că el e resimţit, că e cunoscut, aşadar, de către conştiinţă. Posibilitatea unui inconştient nu s-ar pune absolut deloc atunci când vorbim de sentimente, senzaţii şi afecte”.

(7)În acest context, e cât se poate de important să observăm că funcţionarea corticală e însoţită de conştiinţă numai dacă
este „facilitată” de sistemul reticulat activator din trunchiul cerebral superior. Leziuni de numai doi milimetri pătraţi în această regiune distrug toată conştiinţa. Mulţi cred că asta e din cauză că trunchiul cerebral modulează nivelul cantitativ al conştiinţei sau „starea de veghe”, dar această opinie este neconvingătoare. Conştiinţa generată de trunchiul cerebral superior are un conţinut calitativ propriu. Acesta este afectul. Întrucât conştiinţa corticală depinde de conştiinţa din trunchiul cerebral, rezultă că afectul este forma de bază a conştiinţei. Subiectul capabil să simtă este constituit, concret, de afect.

(8)Afectul este o formă extinsă de homeostază, care este un mecanism biologic elementar ce a apărut în mod natural odată cu autoorganizarea. Sistemele cu autoorganizare supravieţuiesc deoarece ocupă stări limitate şi nu se dispersează. Imperativul supravieţuirii

a dus treptat la evoluţia mecanismelor dinamice complexe aflate la baza intenţionalităţii. Aspect esenţial, ipseitatea sistemelor cu autoorganizare le conferă o perspectivă. Iată de ce are sens să discutăm despre subiectivitatea unui asemenea sistem: devierile de la stările lor viabile sunt consemnate de sistem, pentru sistem, ca nevoi.

(9)Afectul dă valenţe hedonice nevoilor biologice, astfel încât deviaţiile în creştere sau în scădere de la punctele de stabilizare homeostatice (erori de predicţie în creştere sau în scădere) sunt simţite ca neplăcere şi, respectiv, plăcere. Fiecare categorie de nevoi – şi sunt multe şi diverse – are o calitate afectivă proprie şi fiecare declanşează programe de acţiune preconizate să readucă organismul în limitele sale viabile. Aceste stări active – adică răspunsuri intenţionale la stările afective – iau forma reflexelor şi instinctelor înnăscute, care sunt completate gradual de învăţarea din experienţă, conform legii afectului[1]. Simţirea de către un organism a unor fluctuaţii ale propriilor nevoi permite realizarea unor alegeri şi, astfel, susţine supravieţuirea în contexte neprevăzute. Aceasta este funcţia biologică a experienţei.

(10) Nevoile nu pot fi simţite toate odată. Ele sunt prioritizate de un triunghi de decizie din mezencefal, unde nevoile curente (erorile de predicţie reziduale, cuantificate ca energie liberă) ce converg în substanţa cenuşie periapeductală sunt clasificate în legătură cu oportunităţile curente (manifestate sub forma unei

„hărţi a predominanţei” bidimensionale în colicului superiori). Asta declanşează programe de acţiune condiţionată, ce se desfăşoară în contexte preconizate în cadrul unei ierarhii profunde a predicţiilor (modelul generativ al prozencefalului extins). Acţiunile generate de afecte prioritizate sunt voluntare, adică sunt supuse unor alegeri de tip aici-şi-acum, şi nu unor algoritmi prestabiliţi. Asemenea alegeri sunt resimţite în conştiinţa exteroceptivă, care contextualizează afectul. Alegerile se fac pe baza unei cântăriri fluctuante a preciziei (numită şi activare, modulare, amplificare postsinaptică) a semnalelor de eroare primite ce sunt făcute predominante de nevoile prioritizate, în timp ce sunt amortizate în memoria de lucru, cu scopul de a minimiza incertitudinea (a maximiza încrederea) faţă de o predicţie curentă referitoare la felul cum poate fi satisfăcută nevoia. Aceasta este „reconsolidarea”. După cum a spus Freud, „conştiinţa ia naştere acolo unde se sfârseste urma mnezică”.

(11) Alegerile de succes duc la ajustări pe termen lung ale predicţiilor senzorimotoare. Astfel, conştiinţa exteroceptivă este travaliu predictiv în desfăşurare, al cărui scop este să stabilească predicţii tot mai profunde (mai sigure, mai puţin conştiente) referitoare la modul în care se pot satisface nevoile. Această consolidare pe termen lung – şi tranziţia de la sisteme de memorie „declarativă” la cele de memorie „nondeclarativă” – necesită reducerea complexităţii modelului predictiv, pentru a facilita posibilitatea de generalizare. Oamenii

aspiră la automatizare – încredere absolută –, dar nu o pot realiza niciodată în întregime. În măsura în care eşuăm, suportăm sentimente. Întrucât nu facem niciodată predicţii perfecte, pulsiunea implicită (când totul merge bine) este CĂUTAREA – abordarea proactivă a incertitudinii, cu scopul de a o rezolva în avans. Când acest afect este prioritizat, îl resimţim drept curiozitate şi interes faţă de lume.

(12) Acestea sunt mecanismele cauzale ale conştiinţei – în toate manifestările ei, neurologice şi psihologice deopotrivă –, aşa cum arată şi cum se simte. Funcţiile fundamentale pot fi reduse la legi naturale, precum legea lui Friston[2]. Aceste legi stau la baza autoorganizării. Ele pot să explice cum şi de ce resimţim opoziţia proactivă faţă de entropie (adică uitare) în aceeaşi măsură în care alte legi ştiinţifice pot să explice alte lucruri naturale. Conştiinţa face parte din natură şi poate fi

exprimată în termeni matematici.

(13)Toate sistemele conştiente cunoscute sunt vii, dar nu toate sistemele vii sunt conştiente. De asemenea, toate sistemele vii se autovalidează, dar nu toate sistemele care se autovalidează sunt vii. Dacă argumentul prezentat aici este corect, atunci, în principiu, putem construi cu ajutorul ingineriei un sistem care să se autovalideze şi să fie conştient artificial. Conştiinţa poate fi *produsă.* Acest lucru va

împlini cele mai fanteziste visuri ale lui Helmholtz şi ale altor membri ai Societăţii de Fizică din Berlin. Totuşi, trebuie să ne gândim ce ne motivează să-l facem, să acceptăm responsabilitatea colectivă pentru posibilele consecinţele dezastruoase şi să acţionăm cu cea mai mare prudenţă.

Anexă

Activare şi informare

Într-o carte importantă pe tema activării cerebrale, Pfaff (2005, pp. 2-6) comentează:

> Satisfăcând nevoia unei „surse de energie” pentru comportament, activarea explică iniţierea şi persistenţa comportamentului motivat la o mare diversitate de specii [...]. Activarea, alimentând mecanismele pulsiunilor, potenţează comportamentul, iar motivaţii şi stimulente specifice explică de ce un animal face un anumit lucru, şi nu altul [...]. *Dictionary ofEthology* nu numai că evidenţiază activarea în contextul ciclului somn-veghe, ci se referă şi la starea generală de receptivitate a animalului, după cum este indicată de intensitatea stimulării necesare pentru a declanşa o reacţie comportamentală. Activarea „împinge animalul spre disponibilitatea de a acţiona dintr-o stare de inactivitate”. În cazul acţiunii direcţionate, unul dintre fondatorii etologiei, Niko Tinbergen, ar spune că activarea furnizează energia motrică pentru un „tipar de acţiune fixă” ca reacţie la un „declanşator”. Dicţionarul nu evită neurofiziologia, întrucât acoperă şi nivelurile de activare indicate de electroencefalograma (EEG) corticală [...]. Generaţii de behaviorişti au teoretizat şi au confirmat experimental că un concept precum activarea este necesar pentru a explica iniţierea, forţa şi persistenţa răspunsurilor comportamentale. Activarea furnizează forţa fundamentală care face animalele

şi oamenii activi şi receptivi astfel încât să manifeste comportamente instinctive sau învăţate, dirijate către obiecte ale unor scopuri. Forţa unui răspuns învăţat depinde de activare şi de pulsiune. Hebb a considerat starea de activare generalizată fundamentală pentru funcţionarea cognitivă optimă. Duffy merge şi mai departe, invocând conceptul „activării" pentru a
explica o parte semnificativă din comportamentul unui animal.

Analizele lui Pfaff asupra componentei principale sugerează că proporţia comportamentului dintr-o serie amplă de date care poate fi explicat prin „activarea generalizată" este de 3045%.

[Duffy] a anticipat că nişte măsurători cantitative fiziologice sau fizice ar permite o abordare matematică a acestui aspect al ştiinţelor behavioriste [...]. Cannon a propus sistemul nervos autonom drept mecanism necesar prin care activarea pregăteşte animalul pentru acţiune musculară. Întregi teorii ale emoţiei s-au bazat pe activarea comportamentului [...]. Malmo a adunat toate aceste materiale, citând dovezi din EEG-uri şi date fiziologice care sunt concordante cu rezultatele comportamentale în stabilirea activării şi stimulării drept componentele principale ce impulsionează toate mecanismele comportamentale [...]. Aceasta este problema clasică a activării: care este modul în care influenţele interne şi externe trezesc creierul şi comportamentul, la om şi la alte animale, în laborator sau într-un cadru natural, etologic? E important să reformulăm şi să rezolvăm această problemă deoarece e vorba de capacitatea de reacţie la mediu, una dintre cerinţele elementare pentru viaţa regnului animal. De asemenea, e oportun să o reformulăm şi să o rezolvăm pentru că noi instrumente neurobiologice, genetice şi computaţionale oferă abordări ale „stărilor comportamentale" ce nu au mai fost posibile până acum [...]. Explicarea activării ne va permite să înţelegem stările comportamentale ce se află la baza unui număr mare de mecanisme specifice de reacţie. Nu numai că e strategic să analizezi mai multe comportamente odată, ci şi

elucidarea mecanismelor unor stări comportamentale duce la o înţelegere a dispoziţiei şi temperamentului. Altfel spus, o bună parte din neuroştiinţele secolului XX au vizat explicarea particularităţii unor conexiuni specifice stimul-răspuns. Acum avem cum să dezvăluim mecanismele unor clase întregi de răspunsuri sub denumirea de „control al stării". Cele mai importante sunt mecanismele ce determină nivelul de activare [...]. Orice definiţie cu adevărat universală a activării trebuie să fie elementară şi fundamentală, primitivă şi nediferenţiată şi să nu derive din funcţii superioare ale SNC. Nu poate fi limitată de condiţii sau criterii temporare, specifice. De exemplu, nu se poate limita la explicarea răspunsurilor la un singur tip de senzaţie. Trebuie incluse şi activitatea motoare voluntară şi răspunsurile emoţionale. Aşadar, propun următoarea definiţie operaţională care este satisfăcătoare la un nivel intuitiv şi care va duce la măsurători cantitative precise: *„activarea generalizată" este mai mare la un animal sau la un om care este: (S) mai atent la stimuli senzoriali de toate tipurile, (M) mai activ sub aspect motric şi (E) mai reactiv din punct de vedere emoţional*. Aceasta este o definiţie concretă a forţei fundamentale din sistemul nervos [...]. Toate cele trei componente pot fi măsurate cu precizie [...]. În mod clar, există o neuroanatomie a activării generalizate, există neuroni ale căror tipare de descărcare duc la ea, precum şi gene a căror pierdere o perturbă. Prin urmare [...], activarea generalizată este starea comportamentală produsă de căile de activare, de mecanismele electrofiziologice şi de influenţele genetice. Faptul că aceste mecanisme produc aceleaşi atenţie senzorială (S), reactivitate motoare (M) şi reactivitate emoţională (E) ca în definiţia noastră demonstrează existenţa unei funcţii de activare generalizată, precum şi acurateţea definiţiei sale operaţionale.

Pfaff continuă: „Deoarece activarea SNC depinde de surprindere şi de impredictibilitate, cuantificarea ei corectă depinde de matematica *informaţiilor"* (p. 13, sublinierea mea). Ecuaţia lui Shannon (1948) face informaţiile măsurabile, după cum explică Pfaff:

Dacă orice eveniment este perfect regulat – de pildă, bătaia unui metronom –, evenimentul următor (bătaia următoare) nu ne spune nimic nou. Are o probabilitate extrem de ridicată (*p*) de a apărea în exact acel interval de timp (*time bin*) [...]. Nu avem nicio incertitudine că va avea loc bătaia, indiferent de *time bin*. În ecuaţia lui Shannon, informaţiile din orice eveniment sunt invers proporţionale cu probabilitatea lor. Altfel spus, cu cât suntem mai nesiguri în privinţa apariţiei evenimentului respectiv, cu atât mai multe informaţii se transmit, în mod inerent, când acesta se întâmplă [...]. Când toate evenimentele dintr-o serie de evenimente sunt la fel de probabile, informaţiile au valoarea maximă. Dezordinea maximizează fluxul de informaţii. Termenul tehnic, împrumutat din termodinamică, pentru dezordine în ecuaţia lui Shannon este *entropie*. Simbolul lui Shannon pentru entropie este *H* [...]. Conţinutul de informaţie inerent într-un eveniment *x* oarecare este:

$$H(x)=_{j}p(jc)\log_{2}\text{-h-}pW$$

unde *p*(*x*) este probabilitatea evenimentului *x*.

Pfaff rezumă (pp. 19-20):

Pentru ca un animal inferior sau un om să fie activat, trebuie să existe schimbare în mediul său [interoceptiv sau exteroceptiv]. Dacă există schimbare, trebuie să existe incertitudine cu privire la starea mediului. Din punct de vedere cantitativ, în măsura în care există incertitudine, predictibilitatea scade. Având în vedere aceste consideraţii, putem folosi [ecuaţia lui Shannon] ca să afirmăm că, pe cât mediul este mai puţin predictibil şi entropia este mai mare, pe atât există mai multe informaţii. Activarea creierului şi comportamentului, precum şi calculele informaţionale sunt unite inseparabil.

Pe scurt, stimulii necunoscuţi, nepreconizaţi, dezordonaţi şi

neobişnuiţi (cu multe informaţii) produc şi susţin răspunsuri de activare (p. 23).

> Teoria informaţiei s-a ascuns în tot acest timp în spatele unor cercetări comportamentale şi date neurofiziologice. Mai întâi, conform unei logici simple şi clare, gândeşte-te la ce este necesar pentru ca un animal sau un om să se hotărască să acţioneze. În al doilea rând, gândeşte-te la ce este necesar ca să recunoşti un stimul familiar (obişnuinţa) şi ca să acorzi o atenţie specială unui stimul nou. În al treilea rând, din punctul de vedere al experimentatorului, teoria informaţiei furnizează metode pentru calcularea conţinutului semnificativ al aşa- numitelor *spike trains* şi pentru cuantificarea încărcăturii cognitive a anumitor situaţii de mediu. Se pot pune întrebări noi: de câtă distorsiune a unui câmp de stimuli senzoriali e nevoie pentru noutate? Ce tipuri de generalizare din partea unui anumit tip de stimul sunt permise pentru un anumit tip de răspuns? Abordarea teoriei informaţiei ne va ajuta să transformăm combinaţia dintre genetică, neurofiziologie şi comportament într-o ştiinţă cantitativă. Putem folosi „matematica activării” ca să ne ajute în analiza mecanismelor neurobiologice.

În cele din urmă, Pfaff concluzionează (pp. 138-145):

> Sistemele de activare ale SNC se luptă eroic cu legea a doua a termodinamicii într- un mod foarte special. Ele răspund selectiv la situaţii de mediu ce au o entropie inerent mare – un grad ridicat de incertitudine şi, prin urmare, de conţinut informaţional. Dar, când răspund, sistemele de activare ale SNC reduc eficient entropia prin comprimarea tuturor acelor informaţii într-un singur răspuns legitim [...]. Neurobiologia activării este neuroştiinţa schimbării, incertitudinii, impredictibilităţii şi surprizei – adică a ştiinţei informaţiei. Pe parcursul tuturor analizelor de până acum asupra mecanismelor activării din SNC – neuroanatomice, fiziologice, genetice şi comportamentale –, conceptele teoriei informaţiei se dovedesc a fi utile. Matematica informaţiei furnizează căi pentru clasificarea răspunsurilor la stimulii naturali. De fapt, celulele nervoase codifică probabilităţi şi incertitudini, rezultând că ele pot să ghideze comportamentul

în circumstanţe nepredictibile. Însăşi activarea SNC depinde cu totul de schimbare, incertitudine, impredictibilitate şi surpriză. Fenomenul numit obişnuinţă, o scădere a amplitudinii răspunsului la repetarea aceluiaşi stimul, se întâlneşte peste tot în neurofiziologie, ştiinţele comportamentale şi fiziologia SNA, arătându- ne în ce mod conţinutul informaţional în scădere duce la scăderea activării SNC. Astfel, teoria activării şi teoria informaţiei au fost făcute una pentru cealaltă.

E important să recunoaştem că „matematica informaţiei" explică comportamentul neuronilor atât în procesele de activare, cât şi în cele de învăţare, ce, combinate, determină ceea ce *face* creierul. Aşadar, deşi „informaţia" nu este un construct fiziologic, ea explică în mod legitim activitatea fiziologică a creierului. Ea e *funcţia* selectată de evoluţie, iar fenotipurile fiziologice vin după.

Mulţumiri

Le sunt recunoscător următorilor prieteni şi colegi pentru că au citit mai multe versiuni ale capitolelor: Richard Astor, Nikolai Axmacher, Samantha Brooks, Aimee Dollman, George Ellis, Karl Friston, Eliza Kentridge (care este mult mai mult decât o prietenă), Joe Krikler, Joshua Martin, Lois Oppenheim, Jonathan Shock, Pippa Skotnes şi Dawie van den Heever. Îi sunt deosebit de îndatorat lui Ed Lake pentru că a făcut manuscrisul mult mai accesibil lecturii – nu am mai văzut redactor care să lucreze atât de temeinic. Trevor Horwood a făcut corectura manuscrisului, cu o contribuţie din partea lui Tim James.

Aş vrea să-i mulţumesc şi lui sir Sydney Kentridge pentru că mi-a permis să-i folosesc casa din Chailey, unde am scris cea mai mare parte a cărţii în iernile din 2018-2019 şi 2019-2020. Ca de fiecare dată, în culise au

fost bravele mele asistente, Paula Barkay şi Eleni Pantelis. Această carte, ca mai tot ce am făcut până acum, nu s-ar fi întâmplat fără ele.

Note

Capitolul 1. Materia viselor

1. Popper (1963). E adevărat că nu toată lumea este de acord cu
   această formulare despre cum funcţionează ştiinţa. Cu toate acestea, aproape toţi specialiştii în ştiinţe naturale subscriu la ea. *Acum, că citeşti aceste note, se impune o explicaţie. Notele se adresează în principal cititorilor din mediul academic care sunt interesaţi de (sau sunt familiarizaţi cu) diversele lucrări de specialitate cu care se înrudeşte această carte. Notele pot fi ignorate fără probleme de publicul larg – principalii cititori cărora mă adresez*.
2. Freud (1893a), p. 13.
3. Sacks (1984), p. 164.
4. Freud (1895), p. 160. După vreo 36 de ani, Freud i-a scris într-
   un stil emoţionant lui Albert Einstein despre lipsa prestigiului ştiinţific al psihologiei prin comparaţie cu fizica (Freud, 1994, p. 239): „Într-adevăr, nu e cu totul regretabil să optezi pentru psihologie. Nu există o temă mai măreaţă, mai bogată şi mai misterioasă,

demnă de fiecare efort al intelectului uman, decât viaţa minţii. Cu siguranţă, psihologia este cea mai frumoasă dintre toate doamnele nobile, doar că prinţul ei e condamnat să rămână nefericit în dragoste”.

5. *„Ador* acest pasaj din Freud – sunt foarte fericit că l-ai găsit. După cum spui, cu generozitate, ceva analog se poate spune şi despre propriile mele istorice de caz şi despre istoricele unor cazuri neurologice (cel puţin neuropsihologice) în general. L-am citat (deşi nu ştiu dacă va supravieţui – manuscrisul meu a devenit mult prea lung şi cu prea multe note de subsol) într-un text tocmai finalizat, mai degrabă general – «Scotoma» –, despre uitare şi neglijare în ştiinţă.” (Scrisoare de la Sacks, datată 2 ianuarie 1995)

6. După cum a scris la vremea respectivă Semir Zecki, specialist în neuroştiinţe: „Majoritatea dintre noi s-ar îngrozi la gândul de a investiga ceea ce pare să fie o problemă atât de impenetrabilă” (Zeki, 1993, p. 343).

7. Aserinsky şi Kleitman (1953).

8. Dement şi Kleitman (1957).

9. Vezi Freud (1912), p. 259: „Există un produs psihic care se poate întâlni la persoanele cele mai normale şi care oferă totuşi o analogie extrem de izbitoare cu manifestările cele mai sălbatice ale nebuniei, produs ce nu a fost mai inteligibil pentru filosofi ca nebunia însăşi. Mă refer la vise. Psihanaliza se bazează pe analiza visului; interpretarea visului este partea cea

mai completă a muncii pe care a realizat-o până astăzi tânăra ştiinţă” *[Opere esenţiale,* vol. III: *Psihologia inconştientului,* „Câteva observaţii asupra conceptului de inconştient în psihanaliză”, traducere din limba germană de G. Lepădatu, Trei, Bucureşti, e-book].

10. Popper (1963).
11. Din punct de vedere strict anatomic, talamusul nu este considerat parte a trunchiului cerebral. Totuşi, din punct de vedere fiziologic, unii dintre nucleii săi „nonspecifici” fac parte din sistemul reticulat activator; prin urmare, sunt grupaţi alături de funcţiile trunchiului cerebral (de unde şi sintagma „sistem activator reticular-talamic extins”, ERTAS). Nucleii talamici „specifici”, care acţionează în principal ca staţii de transmitere a semnalelor senzoriale, sunt grupaţi împreună cu funcţiile cortexului. În această carte voi folosi termenii „trunchi cerebral” şi „cortex” în principal pentru a desemna deosebirea funcţional-anatomică dintre *stimularea* ERTAS şi, respectiv, *reprezentarea* talamo-corticală. De aceea, voi considera structuri ale trunchiului cerebral nu numai talamusul şi hipotalamusul „nonspecifici”, ci şi nucleii prozencefalului bazal. Pentru o viziune contemporană asupra sistemului activator reticular-talamic *extins,* vezi Edlow *et al.* (2021).
12. Jouvet (1965).
13. Hobson, McCarley şi Wyzinski (1975).
14. McCarley şi Hobson (1977), p. 1346.
15. *Ibidem,* p. 1219.
16. Singurul pacient care a spus că nu mai visa a suferit

aproape sigur leziuni semnificative mult dincolo de tegmentul mezopontin, care generează somnul REM (din cauza unei hemoragii subarahnoidiene traumatice; Lavie *et al.,* 1984). Prin urmare, modificările la nivelul capacităţii sale de a visa erau dificil de corelat cu o anumită arie cerebrală.

17. Vezi Solms (1997a) pentru o descriere completă a concluziilor mele. Mi-am depus teza în 1991, dar am publicat-o abia după şase ani.
18. Solms (2000a). Principiul „dublei disocieri" în neuropsihologie ne permite să diferenţiem funcţiile mintale la punctele lor de întâlnire: dacă o leziune în aria *X* (a creierului) cauzează pierderea funcţiei *A,* dar nu şi a funcţiei *B,* iar o leziune în aria *Y* cauzează pierderea funcţiei *B,* dar nu şi a funcţiei *A*, atunci funcţia *A* şi funcţia *B* nu pot fi acelaşi lucru. Cu alte cuvinte, în acest caz, funcţia somnului REM şi funcţia visului nu pot fi identice. Ele *se corelează* reciproc (adică apar în acelaşi timp), dar nu sunt acelaşi lucru.
19. Multe alte dovezi, pe lângă constatările mele despre leziuni, susţin această concluzie. De exemplu, există o şansă de 50% de a obţine relatări ale unor vise în primele câteva minute ale somnului (în faza 2 descendentă), cu mult înainte de primul episod REM. De asemenea, visele care nu se pot deosebi deloc de cele REM apar în somnul non-REM cu o frecvenţă în creştere în faza matinală de trezire a ritmului diurn, ceea ce se numeşte „efectul dimineţii târzii". De asemenea, deşi visul este mult mai frecvent în somnul REM decât în somnul non-REM, somnul non-

REM este dominant în comparaţie cu somnul REM. În consecinţă, cel puţin un sfert dintre toate visele apar în timpul somnului non-REM. Vezi Solms (2000a) pentru detalii.

20. Solms (1991, 1995).
21. Frank (1946, 1950), Partridge (1950).
22. Schindler (1953).
23. Hartmann *et al.* (1980).
24. Sharf *et al.* (1978). Studiile ulterioare au arătat că antagoniştii dopaminei au efectul opus (Yu, 2007).
25. Dahan *et al.* (2007).
26. Léna *et al.* (2005).
27. Solms (2011).
28. Solms (2001).
29. Acesta denumire nu e bună, e o revenire la vremurile behavioriste. Există multe tipuri diferite de „recompensă” (adică plăcere) în creier.
30. Rolls (2014), Berridge (2003), Panksepp (1998).
31. Panksepp (1998), p. 155.
32. Activitatea de CĂUTARE dopaminergică (spre deosebire de alte activităţi ale monoaminelor) continuă după instalarea somnului şi este la nivel maxim în timpul somnului REM. Probabil nu e o coincidenţă că coincide cu mişcările sacadate rapide ale ochilor. Mişcările ochilor la oameni, la fel ca adulmecatul şi mişcările mustăţilor la rozătoare, sunt un bun echivalent al activării CĂUTĂRII (vezi Panksepp, 1998).
33. Pace-Schott şi Hobson (1998).
34. Braun (1999), p. 196.
35. *Ibidem,* p. 201.

Capitolul 2. Înainte de Freud şi după

1. Atunci am fost acceptat la specializare, dar nu am început-o până în 1989.
2. De fapt, Freud nu a pus un titlu acestui manuscris nepublicat;
   titlul a fost inventat de traducători. În corespondenţa cu Wilhelm Fliess, Freud l-a numit „Psihologie pentru neurologi”, „Schiţă de psihologie” şi „Psihologia”.
3. Scrisoare către Hallman (1842), publicată în du Bois-Reymond (1918), p. 108. De asemenea, citată frecvent este prefaţa lui du Bois-Reymond la lucrarea sa *Über die Lebenskraft* (1848-1884, pp. xliii-xliv): „Nu există forţe noi ce funcţionează în organisme şi în particulele lor, alte forţe care să nu funcţioneze şi în afara lor. De asemenea, nu există nici forţe care merită numele «forţă vitală». Separarea dintre aşa-numitele naturi organică şi anorganică este complet arbitrară”.
4. Bechtel şi Richardson (1998).
5. Vezi comentariile mele despre acest termen în Solms (în curs
   de apariţie).
6. Freud (1950b), p. 295.
7. Prioritatea lui Freud în formularea poziţiei „funcţionaliste” nu este larg recunoscută (Freud, 1900, p. 536): „[Încercăm să înţelegem] gradul de complicaţie a activităţii psihice, prin descompunerea ei, şi indicarea funcţionării individuale a părţilor individuale ale aparatului. După câte ştiu, nu s-a încercat încă ghicirea alcătuirii instrumentului psihic dintro astfel de descompunere. Mie mi se pare

inofensivă” [Sigmund Freud, *Opere esenţiale,* vol. 2: *Interpretarea viselor,* traducere din limba germană de Roxana Melnicu. Bucureşti: Trei, 2010, pp. 614-615]. *Cf* Shallice (1988).

8. Când Freud a introdus pentru prima dată acest termen ciudat, a explicat că se referă la un nivel al explicaţiei ce încorporează atât psihologia, cât şi biologia (scrisoare către Fliess din 10 martie 1989; Freud, 1950a). După cum i-a scris o dată Freud lui Georg Groddeck: „Inconştientul este îndelung căutata verigă lipsă dintre fizic şi mintal” (scrisoare din 5 iunie 1917). Vezi Solms (2000b). Vezi şi prezentarea mea de la o reuniune a Academiei de Ştiinţe din New York organizată pentru a sărbători centenarul „Proiectului” lui Freud (Solms, 1998). Marele Karl Pribram a luat cuvântul la reuniunea respectivă. Tot acolo l-am întâlnit şi pe Joseph Bogen, un neurofiziolog inovator. Mi-l amintesc clar afirmând cu lejeritate că conştiinţa este generată de nucleii intralaminari ai talamusului, aceasta fiind prima dată când am auzit pe cineva sugerând că cortexul nu este conştient în mod intrinsec. Vezi Bogen (1995).
9. Critica adusă de Freud (1891) localizaţionismului a pus bazele
   abordării „sistemelor funcţionale” care a dominat mai târziu neuropsihologia şi apoi cognitivismul. Am discutat în detaliu despre aceste lucruri în Solms şi Saling (1986) şi Solms (2000b).
10. De unde şi titlul cărţii lui Sulloway (1979), *Freud: Biologist of the Mind.*

11. Freud (1914), p. 78.
12. Freud (1920), p. 83.
13. Scrisoare către Fliess, 20 octombrie 1895.
14. Scrisoare către Fliess, 29 noiembrie 1895.
15. Scrisoare către Fliess, 25 mai 1895.
16. În prezent pregătesc o traducere în engleză a operelor complete ale lui Sigmund Freud legate de neuroştiinţe (în patru volume). Vezi şi ediţia revăzută de mine a traducerilor lui Strachey, *The Revised Standard Edition of the Complete Psychological Works of Sigmund Freud* (24 de volume).
17. Freud (1950b), pp. 303, 316. Cu mulţi ani înaintea lui Freud, Baruch Spinoza a scris că „dorinţa este natura însăşi sau esenţa unei persoane”.
18. Freud (1901), p. 259.
19. Freud (1920), p. 60.
20. Freud (1915a), pp. 121-122, sublinierea mea.
21. Scrisoare către Fliess, 25 mai 1895.
22. Freud (1940), p. 197.
23. Această figură este adaptată după Braun *et al.* (1997). Studiul lui Braun a fost pur descriptiv. Constatările lui sunt compatibile cu teoria lui Freud, dar nu o confirmă experimental deoarece nu au testat nicio predicţie derivată din ea. Totuşi, o studentă a mea (Catherine Cameron-Dow, 2012) a testat recent în mod direct teoria lui Freud că visele protejează somnul. Confirmarea de către ea a ipotezei este acum urmărită într-un studiu mai amplu al colegei mele Tamara Fischmann, în Berlin.
24. Întâmplător, dezbaterea a fost prezidată de nimeni altul decât David Chalmers. Rezultatul a inversat un

vot din 1978 ce urmase prezentării de către Hobson a teoriei sale despre „activare-sinteză” în plenul Asociaţiei Americane de Psihiatrie.

25. Solms şi Saling (1986).
26. Vezi Braun (1999).
27. Malcolm-Smith *et al.* (2012).
28. Solms (2000c), Solms şi Zellner (2012).
29. E interesant că Freud şi-a elaborat conceptul de „pulsiune libidinală” într-o perioadă în care consuma regulat cocaină, un alcaloid ce activează intens sistemul CĂUTĂRII dopaminergic. Oare e prea fantezist să presupunem că experienţa personală a lui Freud cu efectele motivaţionale generalizate ale cocainei a contribuit la faptul că a recunoscut existenţa în minte a unui asemenea mecanism motivaţional universal?
30. În capitolul 7 voi arăta legătura dintre „împlinirea dorinţelor” şi codificarea predictivă şi cea dintre „testarea realităţii” şi ceea ce numim astăzi „eroare de predicţie” (sau eroare de predicţie modulată de precizie).
31. Fotopoulou, Solms şi Turnbull (2004).
32. Turnbull, Jenkins şi Rowley (2004).
33. Fotopoulou şi Conway (2004), Turnbull, Berry şi Evans (2004), Fotopoulou *et al.* (2007, 2008a, 2008b), Turnbull şi Solms (2007), Fotopoulou, Conway şi Solms (2007), Fotopoulou (2008, 2009, 2010a, 2010b), Coltheart şi Turner (2009), Cole *et al.* (2014), Besharati, Fotopoulou şi Kopelman (2014), Kopelman, Bajo şi Fotopoulou (2015).
34. Vezi Turnbull, Fotopoulou şi Solms (2014) pentru

recenzii. Vezi şi Besharati *et al.* (2014, 2016).

35. Zellner *et al.* (2011).
36. Solms şi Turnbull (2002, 2011), Panksepp şi Solms (2012), Solms (2015a).
37. În cele din urmă, aceste prezentări au fost adunate într-un volum: Kaplan-Solms şi Solms (2000).
38. Kandel (1998).
39. Kandel (1999), p. 505.
40. Printre specialiştii în neuroştiinţe se numărau Allen Braun, Jason Brown, Antonio Damasio, Vittorio Gallese, Nicholas Humphrey, Eric Kandel, Marcel Kinsbourne, Joseph LeDoux, Rodolfo Llinăs, Georg Northoff, Jaak Panksepp, Michael Posner, Vilanayur Ramachandran, Oliver Sacks, Todd Sacktor, Daniel Schacter, Carlo Semenza, Tim Shallice, Wolf Singer şi Max Velmans. Printre psihanalişti – Peter Fonagy, Andre Green, Ilse Grubrich-Simitis, Otto Kernberg, Marianne Leuzinger-Bohleber, Arnold Modell, Barry Opatow, Allan Schore, Theodore Shapiro, Riccardo Steiner şi Daniel
    Widlocher.
41. Damasio, Damasio şi Tranel (2013).
42. Freud (1940), p. 198. Pe tot parcursul acestei cărţi, folosesc versiunile revăzute ale traducerilor lui James Strachey (vezi Solms, în curs de apariţie).
43. Freud (1920), p. 24, sublinierea mea.
44. Freud (1940), pp. 161-162, sublinierea mea.
45. Vezi Freud (1923), p. 26: „Eul este înainte de toate o entitate corporală, nu doar una situată la suprafaţă, ci însăşi proiecţia unei suprafeţe. Pentru a ne servi de o analogie anatomică, el poate fi comparat cu

«manechinul cerebral» al anatomiştilor, situat în scoarţa cerebrală: cu capul în jos şi picioarele în sus, ochii întorşi spre înăuntru şi având zona limbajului în stânga [...]. În acest fapt avem o nouă demonstraţie a ceea ce afirmam mai sus în legătură cu Eul conştient, şi anume că el reprezintă în primul rând un Eu corporal" [Sigmund Freud, „Eul şi Se-ul", traducere de G. Purdea şi Vasile Dem. Zamfirescu, în *Opere esenţiale,* vol. 3: *Psihologia inconştientului.* Bucureşti: Trei, e-book].

46. Vezi Solms (2013).

Capitolul 3. Eroarea privitoare la cortex

1. Merker (2007), p. 79. Aceste observaţii au confirmat un raport anterior întocmit de Shewmon, Holmes şi Byrne (1999).
2. Relatarea ce urmează despre constatările lui Merker este parafrazată după raportul pe care l-a publicat (2007).
3. Damasio şi Carvalho (2013), p. 147.
4. Aceste observaţii parafrazează rezumatul lui Merker asupra literaturii de specialitate (2007), p. 74.
5. Panksepp (1998).
6. Vezi Weiskrantz (2009).
7. Urmăreşte videoclipul:

https://blogs.scientificamerican.com/observations/blindsight-seeing-without-knowing-it/

8. Acelaşi lucru e valabil şi în cazul căii de la ochi la corpul geniculat lateral. Vezi figura 6.
9. Zeman (2001).
10. Coenen (2007), p. 88, sublinierea mea.
11. Nu există *răspuns ocular* = 1 punct, deschide la stimul de durere = 2 puncte, deschide la ordin verbal = 3, deschide spontan = 4; nu există *răspuns verbal* = 1, răspuns cu sunete neinteligibile = 2, răspuns cu cuvinte fără legătură = 3, răspuns confuz = 4, răspuns orientat = 5; nu există *răspuns motor* = 1, extensie (decerebrare) = 2, flexie anormală (decorticare) = 3, flexie normală (răspuns de retracţie) = 4, localizarea durerii = 5, răspunde la comenzi verbale = 6.
12. Se afirmă şi că iluzia ia naştere din „eroarea moralistă”. Heather Berlin, specialistă în neuroştiinţe cognitive, a calificat drept „arbitrar” să presupui că abilitatea de a răspunde la stimuli externi implică conştiinţă internă (Berlin, 2013, pp. 25-26): „Ipoteza principală a lui Solms conform căreia copiii cu hidroanencefalie sunt conştienţi este neîntemeiată. Nu putem presupune că a avea un ciclu de somn-veghe şi expresii ale emoţiilor (râs, furie etc.) reclamă conştiinţă […]. Deşi este adevărat că ei pot fi de fapt conştienţi, nu putem presupune că şi sunt. Procesele inconştiente pot fi destul de sofisticate şi de complexe (Berlin, 2011). Esenţa teoriei lui Solms se întemeiază pe o proiecţie a existenţei unei conştiinţe bazate pe ceea ce par a fi comportamente

emoţionale semnificative, un exemplu de «eroare moralistă» (să afirmi că ceva trebuie să fie adevărat pentru că ne face să ne simţim bine să credem afirmaţia respectivă). Oamenii doresc în mod natural să presupună că există conştiinţă”.

Iată ce am scris eu în replică (Solms, 2013, pp. 80-81): „De ce ar trebui să presupunem că manifestările emoţionale adecvate contextual, care sunt imediat evocate de stimularea unei anumite regiuni cerebrale şi anulate de leziuni ale aceleiaşi regiuni cerebrale şi care corespund trăirilor noastre afective, *nu* corespund unor trăiri afective ale acestor copii şi animale? Cu siguranţă că această presupunere ar fi mai «arbitrară» decât a mea. Singura dovadă în acest sens este că aceşti copii şi animale nu îşi pot «declara» sentimentele *în cuvinte”.*

13. Cunoscute sub denumirea de legile de asociaţie: prin contiguitate, repetiţie, atenţie, similaritate etc.
14. Apercepţia este „procesul prin care experienţa nouă este asimilată şi transformată de restul experienţelor trecute ale unei persoane pentru a forma un nou întreg” (Runes, 1972).
15. Meynert (1867).
16. Această afirmaţie omite în mod convenabil faptul că creierul este mai mare la alte mamifere (precum elefanţii). Cortexul uman nu este mai mare decât cel al altor mamifere nici măcar în privinţa proporţiei dintre cortex şi mărimea corpului sau dintre cortex şi formaţiunile subcorticale.

*U_*. Campbell (1904), pp. 651-652.

18. Meynert (1884; traducere în engleză, 1885), p. 160.

19. Munk (1878, 1881).
20. S-a considerat că baza anatomică pentru deosebirea dintre orbire şi cecitatea minţii este faptul, descoperit de Paul Flechsig (1901, 1905), că cortexul striat „de proiecţie” conţine celule primordiale conectate direct la periferia retiniană. Aceste celule primordiale sunt mielinizate la naştere şi, prin urmare, nu conţin imagini *mnezice.* Cortexul „de asociere” din jur – vehiculul tuturor funcţiilor psihice – se mielinizează mult mai târziu. Afirmaţia e valabilă în cazul celorlalte regiuni corticale specifice inputurilor senzoriale.
21. Wilbrand (1887, 1892). Vezi traducerea făcută de mine în engleză a studiului de caz original al lui Wilbrand, unde multe dintre aceste idei teoretice sunt discutate detaliat (Solms, Kaplan-Solms şi Brown, 1996).
22. Această observaţie a lui Wilbrand pleca de la o observaţie similară anterioară a lui Charcot (1883) despre un pacient care avea vise *nonvizuale,* astfel apărând conceptul de „sindrom Charcot-Wilbrand”: incapacitatea de a revizualiza şi a recunoaşte obiecte vizuale ziua şi pierderea capacităţii de a visa noaptea. Vezi Solms, Kaplan-Solms şi Brown (1996) şi Solms (1997a) pentru o analiză critică a conceptului Charcot-Wilbrand în contextul concluziilor mele ulterioare.
23. În consecinţă, a făcut o distincţie între surditatea comună şi „surditatea verbală” (afazie). Despre afazia lui Wernicke se spunea că este cauzată de leziuni la nivelul regiunii corticale ce conţine imagini mnezice auditive pentru cuvinte – amintirea

sunetelor vorbirii –, iar surditatea comună apărea în urma unor leziuni la nivelul căilor subcorticale ce conectează această regiune de senzaţiile auditive din exterior. Paul Pierre Broca (1861, 1865) descrisese anterior o formă paralelă de afazie, cauzată de afectarea imaginilor motoare pentru cuvinte – adică afectarea programelor deprinse despre modul în care *produci* sunetele vorbirii. Conform unui student al lui Wernicke, Ludwig Lichtheim (1885), alte forme de afazie erau cauzate de afectarea căilor transcorticale „de asociere" ce duc de la imaginile mnezice auditive şi motoare pentru cuvinte la idei *abstracte,* ceea ce conferea imaginilor concrete semnificaţia lor. Heinrich Lissauer (1890) a subdivizat în mod similar „cecitatea psihică" în două tipuri, un tip „aperceptiv", cauzat de afectarea imaginilor mnezice vizuale, şi un tip „asociativ", cauzat de afectarea căilor transcorticale ce duc de la imaginile mnezice pentru obiecte vizuale la cele pentru idei abstracte. Freud (1891) a fost cel care, mai târziu, a redenumit cecitatea psihică drept „agnozie" vizuală. Apoi, Hugo Liepmann (1900) a împărţit în mod similar „paralizia psihică" (apraxia) în tipuri aperceptive şi asociative – apraxie „chinestezică a membrelor" şi, respectiv, „ideomotoare" –, după cum s-a menţionat în text.

24. Liepmann (1990).
25. În prefaţă a scris: „Cititorul nu va găsi o altă definiţie a «psihiatriei» în această carte în afară de cea dată pe pagina de titlu: Tratat clinic de boli ale prozencefalului. Denumirea istorică pentru psihiatrie, adică «tratarea sufletului», implică mai

mult decât putem să realizăm şi transcende graniţele investigării ştiinţifice riguroase”.

26. Vezi Absher şi Benson (1993) şi Goodglass (1986). E interesant că Antonio Damasio i-a fost student lui Geschwind.
27. Se întâmplă că ştim că pacienţii cu afecţiuni corticale şi fără limbaj au conştienţa nealterată, deoarece pot să-şi comunice sentimentele în alte feluri şi o şi fac. Vezi Kaplan-Solms şi

Solms (2000) pentru descrieri detaliate ale unor rapoarte introspective nonverbale făcute de pacienţi afazici de diferite tipuri. După o oarecare ezitare iniţială în secolul al XlX-lea, consensul general a fost întotdeauna că pierderea limbajului nu afectează „inteligenţa” în mod semnificativ. *2^.* Merker (2007), p. 65.

29. *Cf.* distincţia făcută de Ned Block (1995) între conştiinţa „fenomenologică” şi „de acces”.
30. Vezi Craig (2009, 2011).
31. Vezi Dehaene şi Changeux (2005); Baars (1989, 1997). Sintagma „gânduri de ordin superior” este din Rosenthal (2005).
32. Qin *et al.* (2010), Mulert *et al.* (2005).
33. Dialogul este extras dintr-un raport verbal al cazului prezentat la Congresul de Neuropsihanaliză din Berlin în 2011. Studiul de caz a fost publicat ulterior în Damasio, Damasio şi Tranel (2013).
34. Damasio, Damasio şi Tranel (2013).
35. Harlow (1868).
36. LeDoux şi Brown (2017).
37. Vezi LeDoux (1999), p. 46, sublinierea mea: „Atunci

când stimulii electrici aplicaţi pe amigdala umană generează sentimente de frică (vezi Gloor, 1992) nu e din cauză că amigdala «simte» frica, ci din cauză că diversele reţele activate de amigdală furnizează memoriei funcţionale inputuri care sunt *etichetate* drept frică”.

38. Whitty şi Lewin (1957), p. 73.
39. Solms (1997a), p. 186, Cazul 22.
40. Epitetul a fost introdus de Karl Jaspers (1963).

Capitolul 4. Care este obiectul experienţei?

1. Vezi Solms şi Saling (1990) pentru o discuţie detaliată despre
dezacordurile dintre Freud şi Meynert şi alţi specialişti în neuropsihologie de la vremea respectivă. Iniţial, şi-a prezentat opiniile separate într-un manuscris nepublicat (Freud, 1887). Acestea au fost tipărite pentru prima dată în Freud (1888). Apoi le-a reiterat în cartea sa din 1891 şi le-a dezvoltat în Freud (1893b) şi, în cele din urmă, în scrisoarea către Fliess din 6 decembrie 1896.
2. Freud (1891), traducerea mea. Freud (1886), p. 14, a raportat
că a făcut „vizite repetate” la laboratorul lui Munk din Berlin. Traseul căilor neuronale din trunchiul cerebral constituia o preocupare aparte a lui Freud în studiile sale de anatomie, în anii 1880.
3. Freud (1891), traducerea mea.
4. *Ibidem,* traducerea mea. Freud a pus sub semnul întrebării şi
baza anatomică şi fiziologică a distincţiei făcute de

Meynert şi de alţii între etapele aperceptive şi cele asociative ale procesării corticale (*ibidem*).

5. Scrisoare către Wilhelm Fliess, datată 6 decembrie 1896 (Freud, 1950a, p. 233).
6. Prin urmare, la Freud, sistemele inconştiente şi preconştiente

   coincid cu ceea ce numim astăzi memorie pe termen lung „nondeclarativă” şi „declarativă”, iar sistemul conştient coincide cu ceea ce numim memorie pe termen scurt. Prima şi ultima dintre cele cinci etape ale lui Freud (percepţia şi conştiinţa) sunt *stări* ale neuronilor, şi nu urme.
7. Kihlstrom (1996).
8. Bargh şi Chartrand (1999), p. 476.
9. Squire (2009).
10. Claparède (1911).
11. Galin (1974), p. 573.
12. Vezi McKeever (1986).
13. Crick şi Koch (1990), Newman şi Baars (1993), Dehaene şi Naccache (2001), Bogen (1995), Edelman (1990), Marc şi Llinas (1994), Tononi (2012).
14. Nagel (1974). Vezi şi Chalmers (1995a) şi Strawson (2006).
15. Freud (1915b), p. 177, explică: „Toată deosebirea [dintre idei şi afecte] provine din faptul că reprezentările sunt investiri – în mod fundamental, ale urmelor mnezice –, în timp ce afectele şi sentimentele corespund proceselor de descărcare, ale căror ultime manifestări sunt percepute ca senzaţii. În situaţia actuală a cunoştinţelor noastre despre afecte şi sentimente, nu putem enunţa mai

clar această deosebire” [Sigmund Freud, *Opere esenţiale,* vol. III: *Psihologia inconştientului,* „Inconştientul”, traducere din limba germană de G. Lepădatu. Bucureşti: Trei, e-book].

16. De exemplu, în *The Quest for Consciousness,* Francis Crick şi Christoph Koch au explicat că afectul implică „aspectele mai dificile ale conştiinţei” (Crick, 2004, p. xiv); aşa că ei au „ignorat în mod intenţionat [...] modul în care emoţiile [...] contribuie la formarea şi modelarea coaliţiilor neuronale care sunt suficiente pentru percepţia conştientă”, axându-se în schimb pe „aspecte ale conştiinţei mai abordabile din punct de vedere experimental”, precum percepţia vizuală (Koch, 2004, p. 94). E uimitor că au considerat că mecanismul neocortical al vederii conştiente este o problemă mai simplă decât mecanismul trunchiului cerebral primitiv al afectului, deşi au recunoscut în acelaşi timp că afectul modelează coaliţiile neuronale care sunt (aşa-zis) suficiente pentru percepţia conştientă. În orice caz, chiar dacă problema afectului este *într-adevăr* „mai dificilă”, acesta nu este un motiv suficient pentru a o lăsa în afara descrierii pe care o facem conştiinţei. Cam cum ar fi să nu includem efectele cuantice într-o descriere a fizicii!
17. Nu vreau să sugerez că le-au neglijat cu totul; vezi Hume (1748), *Cercetare asupra intelectului omenesc.*
18. Acest termen îi poate fi atribuit lui Panksepp (2011).
19. Skinner (1953), p. 160.
20. Thorndike (1911).
21. Mazur (2013).

22. Vezi Leng (2018), capitolele 16-19, pentru o descriere fascinantă a cercetărilor relevante.
23. Vezi Solms şi Panksepp (2010).
24. Panksepp (1974).

Capitolul 5. Sentimentele

1. Alboni (2014).
2. Următorul paragraf îl parafrazează pe Merker (2007).
3. Vezi Nummenmaa *et al.* (2018) – un articol foarte interesant –, care susţine mai multe dintre concluziile la care ajung eu mai jos, de exemplu, cu privire la imposibilitatea existenţei unor afecte „neutre”, la concretizarea afectelor şi la natura categorială a afectelor.
4. De exemplu, frica aduce cu sine respiraţia rapidă, creşterea ritmului cardiac, redirecţionarea sângelui din intestine spre musculatura scheletului şi, astfel, starea de alertă tonică şi tendinţa de a fugi. Dacă sentimentul de frică ar aduce cu sine respiraţia lentă, scăderea ritmului cardiac, redirecţionarea sângelui de la musculatura scheletului spre intestine şi, astfel, un repaus inert, atunci *nu ar fi frică.* Caracterul intrinsec al afectelor este evident în termografiile corporale totale în prezenţa diferitelor emoţii (vezi Nummenmaa *et al.,* 2018). Vezi şi Niedenthal (2007).
5. Discut mai mult despre aceste probleme filosofice în capitolul 11.

6. Urologii numesc această situaţie *latch-key urgency* („urgenţă de tip cheia de la casă”).
7. Ekman *et al.* (1987).
8. Barrett (2017).
9. Vezi Panksepp (1998) şi Panksepp şi Biven (2012) pentru detaliile empirice şi referinţele bibliografice.
10. Vezi LeVay (1993).
11. Pentru cei interesaţi de detaliile anatomice: de obicei, la indivizii de sex masculin, centrul circuitului DORINŢEI SEXUALE este hipotalamusul anterior (mai ales nucleii interstiţiali), de unde coboară prin nucleii pat ai striei terminale spre PAG. Din punct de vedere chimic, hormonul steroid testosteron (eliberat de testicule şi care acţionează în mare măsură asupra hipotalamusului anterior) mediază eliberarea în creier a unei peptide numite vasopresină, care explică excitarea sexuală masculină. De obicei, la indivizii de sex feminin, controlul sexual se localizează în hipotalamusul ventromedial, iar substanţele principale sunt estrogenul şi progesteronul (amândouă eliberate de ovare, echivalentul la indivizii de sex feminin al testiculelor). Aceşti hormoni mediază în creier activitatea oxitocinei, o peptidă care guvernează o bună parte din răspunsul sexual tipic la indivizii de sex feminin. DORINŢA SEXUALĂ este mediată şi de alte peptide, cum ar fi LH-RH şi CCK.
12. CĂUTAREA este mediată şi de neurotransmiţătorul numit glutamat şi de o serie de peptide, precum oxitocina, neurotensina şi orexina.
13. Revin la tema gândirii, cu unele detalii, în capitolul

10.
14. Îşi are originile în părţile mediale ale amigdalei şi trece prin nucleii pat ai striei terminale şi prin hipotalamusul median şi perifornical, în drum spre PAG. Neuromodulatorul său de comandă este o peptidă numită substanţa P, care acţionează împreună cu glutamatul şi acetilcolina. Acest din urmă fapt ar putea explica de ce tulburarea de comportament REM este atât de frecvent exprimată prin FURIE.
15. Circuitul FRICII apare din amigdala centrală şi cea bazolaterală. Din punct de vedere chimic, este mediată de neurotransmiţătorul glutamat, pe lângă peptidele DBI, CRF, CCK, a-MSH şi NPY.
16. Vezi Tranel *et al.* (2006), care se axează explicit pe emotivitatea *subiectivă* a lui S.M.
17. Vezi Blake *et al.* (2019).
18. Acelaşi lucru se aplică şi altor sarcini similare, precum „intuiţia” în sarcina jocurilor de noroc din Iowa (Turnbull *et al.,* 2014). După cum i-am explicat lui Nicholas Humphrey, care a obiectat la faptul că am folosit termenul „instinct” în loc de „ghicire” în cazul vederii oarbe, totul depinde de întrebările pe care le pui (conferinţa Science of Consciousness, Interlaken, iunie 2019).
19. LeDoux (1996).
20. Bowlby (1969).
21. De exemplu, Yovell *et al.* (2016) şi Coenen *et al.* (2019).
22. Solms şi Panksepp (2010).
23. Prin nucleii pat ai striei terminale, aria preoptică şi

talamusul dorsomedial. Pe lângă mecanismele opioidelor descrise în text, PANICA este mediată de neurotransmiţătorul glutamat şi de neuropeptidele oxitocină, prolactină şi CRF.

24. Acesta este probabil motivul pentru care durerea se somatizează atât de frecvent ca durere fizică. Vezi Eisenberger (2012) şi Tossani (2013).
25. GRIJA este mediată şi de dopamină.
26. Circuitul GRIJII coboară din cortexul cingular anterior, prin nucleii pat ai striei terminale, aria preoptică şi aria tegmentală ventrală până la PAG.
27. Forrester *et al.* (2018).
28. Penksepp şi Burgdorf (2003). Vezi www.youtube.com/watch?v= j-admRGFVNM
29. Sper că acesta nu mai este un joc popular. Nu *aprob* faptul că amerindienii au fost dominaţi de fermieri colonişti, dar cu siguranţă au fost.
30. Empatia derivă din atitudinea intenţională sau „teoria minţii” care, desigur, nu este la fel de dezvoltată la toate speciile de mamifere. Prin urmare, dezvoltarea empatiei nu e în niciun caz un proces automat, după cum ar putea sugera teoria „neuronilor-oglindă”. Empatia nu este un reflex, ci o realizare evolutivă (vezi Solms, 2017a).
31. Totuşi, talamusul dorsomedial şi aria parafasciculară par să fie deosebit de importante, ca şi PAG, desigur. Dacă există neurotransmiţători principali ai circuitului JOACĂ, atunci vorbim de opioidele mu, dar asta ar putea pur şi simplu să reflecte faptul că siguranţa (adică nivelul scăzut de PANICĂ/MÂHNIRE) este o precondiţie necesară

pentru JOACĂ. Alţi posibili modulatori ai circuitului JOACĂ sunt glutamatul, acetilcolina şi canabinoidele. Circuitul talamocortical identificat de Zhou şi colaboratorii săi (2017) are legătură numai cu un aspect specific al circuitului JOACĂ, şi anume dominarea. Vezi şi Van der Westhuizen şi Solms (2015), Van der Westhuizen *et al.* (2017).

32. Pellis şi Pellis (2009).

Capitolul 6. Sursa

1. Moruzzi şi Magoun (1949).
2. Fischer *et al.* (2006). Leziunea era în apropierea (chiar deasupra) nucleului parabrahial medial. E interesant că în această regiune Hobson a identificat celulele-sursă colinergice pentru somnul REM. Vezi şi Parvizi şi Damasio (2003), Golaszewski (2016).
3. Blomstedt *et al.* (2008). Electrodul a fost introdus în substanţa neagră. Locul vizat era nucleul subtalamic.
4. Damasio *et al.* (2000), Holstege *et al.* (2003).
5. Mă refer la convexitatea neocortexului, nu la hipocamp.
6. Garcia-Rill (2017).
7. Holeckova *et al.* (2006).
8. Această utilizare a termenilor „canal” şi „stare” i se datorează lui Mesulam (2000). În capitolul 9 folosesc analogia „modurilor de operare” pentru funcţiile „stărilor”. Pentru o descriere interesantă a aspectului concret al

creierului în diferite stări sau moduri de operare, la nivel celular, vezi

Abbott (2020).

9. Panksepp (1998), p. 314. Acest citat din Panksepp continuă mai jos şi include următoarele două note de final.
10. Bailey şi Davis (1942).
11. Depaulis şi Bendler (1991).
12. Vezi Walker (2017), pentru o prezentare foarte bună a modului în care funcţionează acest ritm. Chiar şi moluştele

şi echinodermele (precum stelele-de-mare) prezintă un ciclu somn/veghe. Panksepp (1998), p. 135, arată că reglarea somnului este mai veche din punct de vedere filogenetic decât sistemul reticulat activator. Pornind de la această idee, el face o sugestie contrariantă: „Ceea ce acum este mecanismul somnului REM la origini media activarea selectivă a emotivităţii. Înainte de apariţia unor strategii cognitive complexe, e posibil ca animalele să-şi fi generat cea mai mare parte a comportamentului din rutine psihocomportamentale al proceselor primare pe care acum le recunoaştem drept sisteme emoţionale primitive [...]. Cu alte cuvinte, e posibil ca multe dintre comportamentele animalelor preistorice să fi apărut în mare parte din subrutine emoţionale preprogramate. Aceste soluţii comportamentale simpliste au fost înlocuite în cele din urmă de abordări cognitive mai sofisticate care au necesitat

nu numai mai mult neocortex, ci şi mecanisme de activare noi, pentru a susţine funcţii eficiente în starea de veghe în interiorul acelor zone cerebrale emergente”.

În lumina a ceea ce propun mai jos, sugestia lui Panksepp poate fi reformulată astfel: prozencefalul adaugă la programele motoare instinctuale automate, inferioare o capacitate de a modula *din punct de vedere contextual* comportamentul emoţional – şi, astfel, de a învăţa din experienţă.

13. Dar nu vorbim despre conceptul filosofic de intenţionalitate sau despre „pe cine vizează intenţia”. Panksepp are în vedere ceva precum „voliţiunea”. Totusi, când mă voi referi la conceptul filosofic, mai târziu, va deveni evident că e profund legat de voliţiune.
14. Astfel, în analiza exhaustivă făcută de Pfaff (2005) pe tema activării, pe care o descrie drept „forţa fundamentală din sistemul nervos”, autorul operaţionalizează termenul după cum urmează: „«Activarea generalizată» este mai intensă la un animal sau la un om care este: (S) mai receptiv la stimuli senzoriali de toate tipurile, (M) mai activ sub aspect motor şi (E) mai reactiv din punct de vedere emoţional”. Având în vedere că „activarea” este un element central în cartea de faţă, reproduc un fragment mai lung din Pfaff (2005) în anexa de la p. 313, care oferă şi o punte utilă spre subiectul capitolului următor. Sunt recunoscător că am această ocazie să subliniez importantele cercetări ale lui Donald Pfaff, care (deja la începutul anilor 1990,

când l-am cunoscut) a manifestat o apreciere neobisnuită pentru ideea de „pulsiune” a lui Freud.

15. PAG are proiecţii spre toţi nucleii sursă a neuromodulatorilor din sistemul reticulat activator. Celelalte destinaţii importante din trunchiul cerebral ale proiecţiilor PAG sunt hipotalamusul medial, nucleul cuneat, formaţia reticulată pontină, nucleul solitar, nucleul *gracilis,* nucleul dorsoreticular şi măduva ventrolaterală. Vezi Linnman *et al.* (2012).
16. Venkatraman, Edlow şi Immordino-Yang (2017). Nu îmi place cuvântul „descendent” în acest context, deoarece PAG *integrează* feedback afectiv cerebral superior şi visceral inferior. Ea „coboară” doar în sensul că duce la un output motor. „Centripet” ar fi un cuvânt mai potrivit, care ar putea fi pus în contrast cu o reţea „centrifugă” (deci cu ceea ce Edlow numeşte reţea „modulatoare”). O reţea „centripetă” ar include reţele atât „descendente”, cât şi „ascendente”. Linnman şi colaboratorii săi (2012) descriu în consecinţă PAG drept locul *interacţiunii dintre* sistemul „limbic descendent” şi cel „senzorial ascendent”.
17. Venkatraman, Edlow şi Immordino-Yang (2017).
18. *Ibidem.*
19. Linnman *et al.* (2012), p. 517, sublinierea mea.
20. *Ibidem,* sublinierea mea. Deloc surprinzător, în studiile de imagistică cerebrală umană, PAG pare să aparţină unei reţele „proeminente” (Seeley *et al.,* 2007).
21. Ezra *et al.* (2015), p. 3468, sublinierea mea.
22. Panksepp şi Biven (2012), p. 413, sublinierea mea.

23. Linnman *et al.* (2012), p. 506, sublinierea mea.
24. Panksepp şi Biven (2012).
25. Se obişnuia să fie numită „griul central”.
26. Cele „din spate” sunt PAG laterală şi PAG dorsolaterală. Cea „din faţă” este PAG ventrolaterală. Această clasificare neglijează PAG dorsomedianã.
27. Vezi Venkatraman, Edlow şi Immordino-Yang (2017): „Când este stimulată, această coloană produce vocalizare emoţională, confruntare, agresiune şi activare simpatică, manifestate printr-o creştere a tensiunii arteriale, a ritmului cardiac şi a celui respirator [...]. În interiorul acestei coloane dorsolaterale/laterale există două părţi. Partea rostrală este responsabilă pentru putere/dominare (generând un răspuns de luptă), iar partea caudală invocă frica (generând un răspuns de fugă) cu creşterea fluxului sangvin spre membre”.
28. Coloana din faţă „primeşte semnale de durere somatice şi viscerale «lente, arzătoare» slab localizate şi, dacă este stimulată, produce *coping* pasiv, comportament de bolnav pe termen lung, încremenire cu hiporeactivitate şi o inhibare a refluxului simpatic [...]. În acest fel, probabil că este implicată în emoţiile de fond precum cele care contribuie la dispoziţie” *(ibidem).*
29. Simone Motta şi colegii săi s-au exprimat astfel (Motta, Carobrez şi Canteras, 2017, p. 39): „[PAG] este considerată în general un loc în aval în reţelele neuronale pentru exprimarea unor comportamente diverse, generând răspunsuri stereotipe. Dar tot mai multe dovezi sugerează că PAG poate să exercite o

modulare mai complexă a multor răspunsuri comportamentale şi că poate să funcţioneze ca un centru unic de furnizare a unui ton emoţional primar pentru a influenţa poziţiile prozencefalice care mediază răspunsuri complexe de aversiune şi de năzuinţă”.

30. În PAG are loc frecvent stimularea cerebrală profundă pentru tratamentul durerii cronice, dar nu reduce capacităţile somato-senzoriale corticale.
31. Merker (2007). La rândul lui, Merker le atribuie această perspectivă lui Penfield şi Jasper (1954).
32. Pentru scopurile acestei cărţi, am simplificat termenul tehnic „triunghi de selectare mezodiencefalic”. De asemenea, modul în care folosesc termenul lui Merker se referă mai mult la o *interfaţă* în decizie decât la un *triunghi* (adică o interfaţă între nevoie şi context). După cum vom vedea în capitolul următor, acţiunea („selectarea acţiunii” pentru Merker) şi percepţia („selectarea ţintei” pentru Merker) – care împreună alcătuiesc contextul – sunt feţele aceleiaşi medalii. Iată cum descrie Merker (2007), p. 70, acest lucru: „Oricât s-a extins ulterior telencefalul, chiar până la a îngropa mezodiencefalul sub un neocortex de mamifer ce sa dezvoltat rapid, nu a fost niciodată nevoie de vreun alt aranjament, iar asta din motive cu adevărat fundamentale. Niciun nerv eferent nu îşi are nucleul motor situat mai sus de etajul mezencefalului. Asta înseamnă că prin intersecţia foarte îngustă din trunchiul cerebral dintre mezencefal şi diencefal [...] trec toate informaţiile prin care prozencefalul poate

vreodată să genereze, să controleze sau să influenţeze comportamentul de orice tip”. Merker numeşte această intersecţie „gâtuirea sinencefalică”. Şi adaugă: „Nu e nevoie să ştim mai mult despre creierul vertebratelor decât faptul că motoneuronii rostrali sunt localizaţi *sub* gâtuirea sinencefalică, ca să ştim că conţinutul informaţional total al prozencefalului trebuie să treacă printr-o reducere masivă a datelor în cursul translatării sale în timp real în comportament”. Faptul că decizia privind ceea ce e de făcut în continuare se ia la acest nivel (al trunchiului cerebral) – adică după ce regiunile prozencefalului şi-au prezentat „cererile” – este ilustrat în mod dramatic de exemplul alarmei de sufocare, discutat mai sus: toate consideraţiile cognitive sunt *trecute într-un plan secundar* de senzaţia de foame de aer, care se declanşează la nivelul trunchiului cerebral. Centrii controlului respirator sunt localizaţi în punte şi în bulbul rahidian.

33. Panksepp (1998), p. 312, a descris aranjamentul funcţional al SINELUI, cu un deceniu înaintea lui Merker, după cum urmează: „Straturile mai profunde ale coliculilor constituie o reprezentare motoare elementară a corpului [obiectiv], nu care interacţionează numai cu sistemele vizual, auditiv, vestibular şi somatosenzorial, ci şi cu circuitele emoţionale învecinate din PAG. PAG elaborează o hartă diferită, de tip visceral a corpului [subiectiv] alături de reprezentările neuronale de bază ale sistemelor durerii, fricii, furiei, detresei de separare, sexuale şi de

comportament matern (după cum este rezumat [în capitolul anterior] în această carte). Lângă PAG se află regiunea mezencefalică locomotoare, ce este capabilă să declanseze tipare neuronale care ar trebui să fie un substrat esenţial pentru a iniţia diferite tendinţe de acţiune coerentă”. De asemenea, Damasio şi Carvalho (2013) au propus acest aranjament funcţional general pentru ceea ce au numit „protosine”. Au arătat că alte hărţi senzoriale heteromodale primitive ale corpului sunt furnizate de complexul parabrahial şi de nucleul tractului solitar. Acestia ar putea fi precursorii evolutivi ai funcţiei integratoare îndeplinite de triunghiul de decizie din mezencefal.

34. Următorul citat este din Merker (2007), p. 73: „[Sinele constient] este singur şi localizat în spatele rădăcinii nasului, în interiorul capului. De acolo *se pare* că ne confruntăm cu lumea vizibilă direct printr-o singură deschizătură ciclopică goală aflată în fruntea capului (Hering, 1879; Julesz, 1971). Totusi, în mod evident, e o simplă aparenţă, întrucât, dacă am fi literalmente şi în fapt situaţi în interiorul capului propriu, ar trebui să vedem nu lumea, ci ţesuturile anatomice din craniu atunci când privim. Deschizătura ciclopică este o invenţie neuronală convenabilă prin care lumea vizuală distală este «inserată» printr-o parte lipsă a corpului vizual proxim, care este «fără cap», cum s-ar spune, sau, mai precis, căruia îi lipseşte partea superioară a feţei (vezi Harding, 1961). În schimb, somestezia menţine o continuitate nealterată în această regiune. Deschizătura goală prin care privim lumea trădează

natura simulată a corpului şi a lumii care ne sunt date în conştiinţă”.

35. *Ibidem,* p. 72. Vezi Stoerig şi Barth (2001) pentru o simulare plauzibilă. Ne facem astfel o idee despre lumea senzorimotoare probabilă a copilului cu hidroanencefalie şi a animalului care a suferit o decorticare cerebrală *(cf.* şi vederea oarbă).
36. White *et al.* (2017), Panksepp (1998), p. 311.
37. Merker (2007), p. 72.
38. Hohwy (2013).
39. După cum s-a şi întâmplat, la câţiva ani de la intervenţia chirurgicală, din cauza unei infecţii a căilor respiratorii superioare (altfel uşor tratabilă) care a fost diagnosticată prea târziu.
40. Vezi Friston (2005).
41. Hohwy (2013), Clark (2015).
42. Să ne gândim, de exemplu, la neuronii mediali dorsali pereche ai musculiţei de oţet, *Drosophila.* În mod surprinzător, pare chiar să fi existat nişte precursori primitivi la nematodul *Caenorhabditis elegans* (vezi Bentley *et al.,* 2016; Chew *et al.,* 2018).
43. Următoarea imagine din Freud (1925), p. 231, ne-ar putea ajuta să ne imaginăm aranjamentul funcţional descris şi, în acelaşi timp, ne-ar putea permite să-i înlocuim termenii „metapsihologici” cu unii fiziologici: „Inervaţiile de investiţie sunt trimise şi din nou retrase din interior spre sistemul *P-Cs* [sistemul „conştiinţei perceptive”], prin şocuri periodice rapide. Atât timp cât sistemul este în acest mod investit, el primeşte percepţii însoţite de conştiinţă şi conduce în continuare excitaţia în sistemele mnezice

inconştiente; îndată ce investirea este retrasă, conştiinţa se stinge, activitatea sistemului fiind sistată. Ar fi ca şi cum [sinele] ar întinde, prin intermediului sistemului *P-Cs*, o antenă în lumea exterioară, care apoi este repede retrasă, îndată ce i-a gustat excitaţiile" [Sigmund Freud, *Opere esenţiale,* vol. III: *Psihologia inconştientului,* „Notă asupra «blocnotesului» magic", traducere de G. Lepădatu. Bucureşti: Trei, e-book]. „Investiţia" este o activare modulatoare. Prin urmare, „inervaţiile de investiţie" care pipăie percepţia corticală în această imagine figurativă sunt impulsuri ale activării nucleului cerebral. Am înlocuit termenul „inconştient" al lui Freud cu „sinele" în acest citat ca să evit faptul că el a combinat în mod eronat aceste două sisteme (Solms, 2013).

44. Merker (2007). Panksepp şi Biven (2012), pp. 404-405, folosesc ritmul teta al hipocampului ca ilustrare a modului în care această „monedă comună" ar putea să funcţioneze din punct de vedere fiziologic, având în vedere faptul că hipocampul codifică *contextul:* „Există indicii sugestive în literatura tradiţională din neuroştiinţe ale anumitor tipuri de oscilaţii sincrone relevante din creier, precum ritmurile de 4-7 Hz din hipocamp numite *ritm teta,* care ajută animalele să exploreze lumea (de exemplu, adulmecatul la şobolani), creând astfel amintiri în hipocamp. Ritmul teta este semnătura neuronală foarte caracteristică a hipocampului atunci când prelucrează activ informaţii. Acest ritm este evident mai ales în timpul activării artificiale a sistemului CĂUTĂRII la

şobolani, un sistem emoţional de primă importanţă pentru culegerea informaţiilor când animalele adulmecă şi îşi explorează mediul (Vertes şi Kocsis, 1997). Cu alte cuvinte, ritmul adulmecării corespunde de obicei frecvenţei curente a undelor teta din hipocamp [...]. Acest lucru ar putea evidenţia modul în care cunoaşterea cognitivă apare din activările modelate ale proceselor afective".

45. „Pipăit" este termenul colocvial pentru senzaţia somatică, ce conţine submodalităţi precum propriocepţia, detectarea temperaturii, detectarea vibraţiei – ca şi alte modalităţi perceptive.
46. Când sunt predominante.
47. Solms (2013).
48. Se poate spune că acetilcolina modulează încrederea în semnalele de eroare, dar această afirmaţie este o simplificare exagerată. Generalizări la fel de exagerat de mari se pot face şi cu privire la serotonină în legătură cu semnalele predictive, la dopamină în legătură cu stările active şi la noradrenalină în legătură cu stările senzoriale. Vezi Parr şi Friston (2018), pentru o opinie mai elaborată.
49. „Fizică" (0uoiKq) înseamnă „cunoaşterea naturii" – a întregii naturi, nu numai a lucrurilor pe care poţi să le vezi şi să le atingi. Ea oferă explicaţii fundamentale ale fenomenelor naturale. Mulţi presupun că fizica studiază doar materia – excluzând astfel mintea prin definiţie –, dar asta ar însemna că mintea nu face parte din natură, ceea ce nu poate să nu ducă la însăşi tema cărţii de faţă. „Un fenomen natural e complet explicat din punct de vedere fizic

numai după ce i s-au descoperit originile până la forţele supreme ale naturii în care se întemeiază şi care sunt active în el.” (Helmholtz, 1892) Se dovedeşte că materia este o stare a energiei (de unde şi $E = MC^2$). „Forţele primare” *explică* fenomenele superficiale, ele nu sunt observate direct, ci sunt deduse. Din acest motiv, ele sunt raportate ştiinţific în termeni nonfenomenologici, ca abstracţiuni.

Cf. Critica raţiunii pure a lui Kant: „Deci numai prin faptul că supunem succesiunea fenomenelor, prin urmare orice schimbare, legii cauzalităţii este posibilă experienţa însăşi, adică o cunoaştere empirică a acestor fenomene; prin urmare, ele însele nu sunt posibile, ca obiecte ale experienţei, decât după această lege” [Immanuel Kant, *Critica raţiunii pure,* ed. a IlI-a îngrijită de Ilie Pârvu, traducere de Nicolae Bagdasar şi Elena Moisuc. Bucureşti: IRI, 1998, p. 113]. Abstracţiunile matematice sunt preferabile din punct de vedere convenţional celor verbale, întrucât cer ca forţele deduse (şi relaţiile dintre ele) să fie măsurabile şi cuantificabile. Acest lucru furnizează o monedă numerică comună în spatele diverselor aparenţe fenomenologice prin care pot fi calculate relaţiile legitime dintre „forţele supreme” ale naturii. După cum a spus Galileo Galilei: „Cartea naturii este scrisă în limbajul matematicii”.

Capitolul 7. Principiul energiei libere

1. Cristalele îşi minimizează energia liberă într-un mod simplu, deoarece starea lor constantă de neechilibru tinde spre un punct fix. Adică ele doar se aranjează în tipare compacte şi rămân aşa, chiar şi atunci când sunt uşor perturbate. Situaţia devine mai complexă atunci când elementul de atracţie are o structură itinerantă cu o dinamică de tipul celei căreia i se conformează creierul uman.
2. Pe scurt, indicele Hirsch este numărul lucrărilor publicate de
   cineva care sunt citate (de egalii săi) de mai multe ori decât rangul lor în succesiunea lucrărilor sale publicate citate. Astfel, dacă a patruzecea cea mai citată lucrare publicată a sa este citată de 42 de ori, dar a patruzeci şi una cea mai citată lucrare publicată este citată de doar 39 de ori (adică mai puţin de 41), atunci indicele Hirsch al autorului este 40.
3. La 20 iulie 2020.
4. Carhart-Harris şi Friston (2010).
5. Freud (1894), p. 60.
6. Friston (2005), sublinierea mea. Această afirmaţie e valabilă chiar şi în cazul articolului său scris împreună cu Carhart- Harris, care începe aşa (Carhart-Harris şi Friston, 2010, p. 1265, sublinierea mea): „Descrierile făcute de Freud proceselor primare şi secundare corespund activităţii autoorganizate din sistemele *corticale* ierarhice şi [...] descrierile pe care le face Eului corespund funcţiilor modului implicit, iar schimburile reciproce sistemelor cerebrale subordonate. Această descriere

neurobiologică se bazează pe o perspectivă asupra creierului ca inferenţă ierarhică sau ca aparat Helmholtz. Din această perspectivă, reţele intrinsece la scară mare ocupă niveluri supraordonate ale sistemelor cerebrale ierarhice care încearcă să-şi optimizeze reprezentarea întregului sistem senzorial. Această optimizare este formulată ca minimizând o energie liberă, un proces similar din punct de vedere formal cu abordarea energiei în formulările freudiene".

7. Friston (2013).
8. Solms şi Friston (2018).
9. Parvizi şi Damasio (2003).
10. Vezi Damasio (2018) şi recenzia mea la această lucrare, Solms (2018a).
11. Ideea lui Einstein la care mă refer este natura cuantică a luminii (Einstein, 1905). Stephen Hawkins a replicat: „Nu numai că Dumnezeu e clar că joacă zaruri, ci uneori ne şi bulversează aruncându-le unde nu pot fi văzute. Mulţi oameni de ştiinţă sunt ca Einstein, în sensul că sunt profund ataşaţi emoţional de determinism. Spre deosebire de Einstein, ei acceptă reducerea capacităţii noastre de a prezice, reducere ce vine din teoria cuantică". Se dovedeşte şi că fenomenele conştiinţei sunt predictibile doar *la un nivel probabilistic.*
12. Asta se întâmpla la University of Southern California, în aprilie 2018. Într-o convorbire telefonică din ianuarie 2019, după ce a citit versiunea publicată a articolului nostru, încă apăra „ştiinţa noastră" pe baza căreia conştiinţa era intrinsec

biologică. Totuşi, din fericire, şi-a schimbat opinia de atunci; vezi Man şi Damasio (2019).

13. E interesant că Hermann von Helmholtz (unul dintre elevii lui Johannes Müller care a înfiinţat Societatea de Fizică din Berlin) a jucat un rol important în formularea acestei legi.
14. Această afirmaţie este valabilă în realitate, în limitele termodinamice, dar nu (strict vorbind) şi în teorie. Din punct de vedere teoretic, ar fi mai corect să spunem: conform celei de-a doua legi, este foarte, foarte *improbabil* ca procesele naturale să fie reversibile. Asta deoarece, când rezolvi ecuaţiile din orice sistem, ai nevoie nu numai de legile relevante ale dinamicii, ci şi de *condiţiile initiale* ale sistemului – iar acestea strică simetria ecuaţiilor fundamentale. Prin urmare, dacă am putea începe cu toate bucăţile unei cesti sparte miscându-se una spre alta la viteza corectă şi toate undele sonore miscându-se înapoi spre cească exact aşa cum trebuie şi la fel şi energia care a trecut pe podea când a aterizat ceasca, atunci totul ar putea reveni ca să recreeze ceasca iniţială intactă. Totusi, pentru că nu suntem niciodată în această situaţie, nu vedem niciodată entropia reducându-se în realitate.
15. Cu excepţia unor perioade foarte, foarte scurte.
16. Când la o substanţă se adaugă căldură, moleculele şi atomii vibrează mai repede. Pe măsură ce atomii vibrează mai repede, spaţiul dintre ei creste. Miscarea şi spaţiul dintre particule determină starea materiei din substanţă. Rezultatul final al miscării moleculare sporite este că substanţa se dilată şi

ocupă mai mult spaţiu.

17. Din punct de vedere tehnic, entropia este asociată cu energia risipită, dar nu e acelasi lucru. Entropia nu are dimensiuni, spre deosebire de energie.
18. Din punct de vedere tehnic, entropia e asociată cu numărul de stări echivalente ce corespund unei configuraţii macroscopice date.
19. Fizicianul Alan Lightman (2018), pp. 67-68, formulează excelent: „Dacă divizăm neobosit spaţiul în părţi din ce în ce mai mici, aşa cum a făcut Zenon, căutând cel mai mic element al realităţii, odată ce ajungem la lumea fantasmagorică a lui Planck, spaţiul nu mai are semnificaţie.

    Cel puţin, ceea ce *înţelegem* prin «spaţiu» nu mai are semnificaţie. În loc să răspundem la întrebarea care este cea mai mică unitate a materiei, invalidăm cuvintele folosite pentru a pune întrebarea. Poate că aceasta este calea întregii realităţi supreme, dacă există aşa ceva. Pe măsură ce ne apropiem, pierdem vocabularul”.

Un alt fizician, Carlo Rovelli (2014), p. 167, face o descriere mai prozaică: „Fundalul spaţiului dispare, timpul dispare, particulele clasice dispar, împreună cu câmpurile clasice. Atunci din ce este făcută lumea? Acum răspunsul e simplu: [...] lumea este făcută în întregime din câmpuri cuantice”. Chiar şi incertitudinea cuantică face parte din universul *fizic.*

20. Un „byte” conţine opt biţi. Prin urmare, un „gigabyte” are opt miliarde de biţi. *Viteza* prelucrării informaţiilor este exprimată în „gigahertzi”, un GHz fiind un miliard de biţi pe secundă. Realitatea fizică a

informaţiei este reflectată de faptul că aceste unităţi sunt şi măsurabile, şi cumpărabile (de exemplu, de la furnizori de servicii de internet).

21. Vezi anexa de la p. 313. Aici trebuie să definim cine este nesigur de ce anume. Aceeaşi afirmaţie este valabilă şi în cazul cuvântului „comunică” din fraza de după cea care conţine nota 22: cine comunică cu cine? Te rog să citeşti mai departe.
22. Gosseries *et al.* (2011).
23. Dacă crezi că această afirmaţie combină entropia informaţională cu entropia termodinamică, vezi citatul din Tozzi, Zare şi Benasich în nota 61, mai jos.
24. După cum vei vedea, e crucial faptul că avem nevoi *multiple,* inclusiv CĂUTAREA.
25. Ia în considerare, de exemplu, inseparabilitatea cuantică dintre două particule: o particulă „conţine informaţii” despre cealaltă.
26. Jaynes (1957). Creşterea volumului sistemului descris în exemplul nostru formal de mai sus (gazul din incintă) îi creşte entropia termodinamică deoarece creşte numărul de microstări posibile ale sistemului care corespund valorilor măsurabile ale variabilelor macroscopice ale acestuia, astfel făcând mai încărcată cu informaţii orice descriere completă a stării sale. Mai exact: în cazul distinct ce foloseşte logaritmi în baza doi, entropia termodinamică redusă este egală cu numărul minim de întrebări binare la care trebuie să se răspundă pentru a specifica complet microstarea, dat fiind că cunoaştem macrostarea. Se poate găsi o relaţie directă şi reală din punct de vedere fizic între entropia

termodinamică şi entropia informaţională dacă se atribuie o unitate de măsură fiecărei microstări ce apare per unitate de substanţă omogenă, apoi se calculează entropia termodinamică a acestor unităţi. Prin analiză teoretică sau prin observaţie, microstările vor apărea cu probabilităţi diferite, iar acest lucru va determina entropia informaţională. Asta arată că entropia lui Shannon este o unitate de măsurare statistică valabilă a microstărilor ce *nu are o altă unitate fizică fundamentală decât unităţile de informaţie.*

27. Shannon (1948).
28. De exemplu, „principiul aspectului dublu” formulat de Chalmers, discutat în capitolul 11.
29. Wheeler i-a fost student lui Niels Bohr, care a formulat principiul complementarităţii. Conform acestui principiu, obiectele au proprietăţi complementare ce nu pot fi *observate* sau *măsurate* simultan. Exemple de proprietăţi complementare sunt particula şi unda, poziţia şi momentul, energia şi durata, rotirea pe axe diferite, valoarea unui câmp şi schimbarea sa (într-o anumită poziţie), precum şi inseparabilitatea şi coerenţa.
30. Wheeler a spus „ceva din bit” *(it from bit).* Iată citatul (Wheeler, 1990, p. 5): „Ceva din bit. Cu alte cuvinte, fiecare ceva – fiecare particulă, fiecare câmp de forţă, însuşi continuumul spaţiu-timp – îşi derivă funcţia, semnificaţia, chiar întreaga existenţă – chiar dacă indirect în unele contexte – din răspunsurile obţinute de maşinării la întrebări de tip da sau nu, opţiuni binare, biţi. *Ceva din bit* simbolizează ideea

că fiecare element din lumea fizică are la bază – o bază foarte profundă, în majoritatea cazurilor – o sursă şi o explicaţie imateriale”.

31. *Ibidem.* Cuvintele citate aici urmează imediat citatului de mai sus.
32. Cititorii atenţi poate observă aici că, în sfârşit, mă ocup de întrebarea pe care mi-am pus-o în copilărie, despre primul răsărit de soare, după cum am discutat în introducere.
33. Darwin (1859). Vezi Friston (2013). Rovelli (2014), pp. 225226, oferă o explicaţie lucidă a acestui lucru: „Un organism viu este un sistem ce se reformează continuu, interacţionând neîncetat cu lumea exterioară. Dintre asemenea organisme, continuă să existe numai cele care sunt mai eficiente în acest sens şi, prin urmare, organismele vii manifestă proprietăţi care le dau posibilitatea să supravieţuiască. Din acest motiv, ele sunt interpretabile, iar noi le interpretăm, în termeni de intenţionalitate, în funcţie de [obiectiv şi] scop. Prin urmare, aspectele finalistice ale lumii biologice (aceasta este descoperirea crucială a lui Darwin) sunt rezultatul selecţiei unor forme complexe eficiente în a persista. Dar modul eficient de a continua să existi într-un mediu schimbător constă în a gestiona mai bine corelaţiile cu lumea exterioară, adică informaţiile; în a culege, a stoca, a transmite şi a elabora informaţii. Din acest motiv există ADN, sisteme imunitare, organe senzoriale, sisteme nervoase, creieri complecşi, limbaje, cărţi, biblioteca din Alexandria, calculatoare şi Wikipedia: ele

maximizează eficienţa gestionării informaţiilor – gestionarea corelaţiilor care favorizează supravieţuirea”.

34. Aici nu contează că sistemele cu autoorganizare sunt întotdeauna vii (nu sunt), ci că întotdeauna sistemele vii se autoorganizează. Şi mai important: nu toate sistemele cu autoorganizare sunt conştiente. Despre fizica originilor vieţii, vezi England (2013), pentru o analiză interesantă dintr-o perspectivă similară cu a lui Friston. Teoria „adaptării impulsionate de disipare” formulată de England afirmă că grupurile de atomi care sunt impulsionaţi de surse externe de energie tind să apeleze la aceste surse, aliniindu- se şi rearanjându-se, astfel încât să absoarbă energia mai bine şi să o disipeze sub formă de căldură (adică îşi minimizează entropia în detrimentul mediului). Mai departe, el arată că această tendinţă de disipare este propice autoreplicării: „O modalitate extraordinară de disipare suplimentară constă în a face mai multe exemplare din tine însuţi”. Totuşi, pentru *biologia* originii vieţii, vezi Lane (2015).
35. Ashby (1947). Vezi şi Conant şi Ashby (1970).
36. Friston (2013), p. 6. Pe monitorul calculatorului se vedeau, desigur, *reprezentări* ale subsistemelor, care nu trebuie confundate cu dinamica statistică. Această afirmaţie e valabilă şi în cazul qualia perceptive asociate cu funcţionarea corticală, după cum vom vedea.
37. Andrei Markov (1856-1922) a fost un matematician rus strălucit, ca şi fratele şi fiul lui. A lucrat în principal în domeniul proceselor stocastice şi este celebru astăzi pentru noţiunile de lanţuri Markov şi

procese Markov. A fost rebelul perfect. Descurajarea instituţională a fost atât de puternică, încât nu a primit recunoaşterea academică în timpul vieţii. În 1912, a protestat împotriva excomunicării lui Lev Tolstoi din Biserica Ortodoxă Rusă, cerând să fie şi excomunicat. Biserica a dat curs cererii sale.

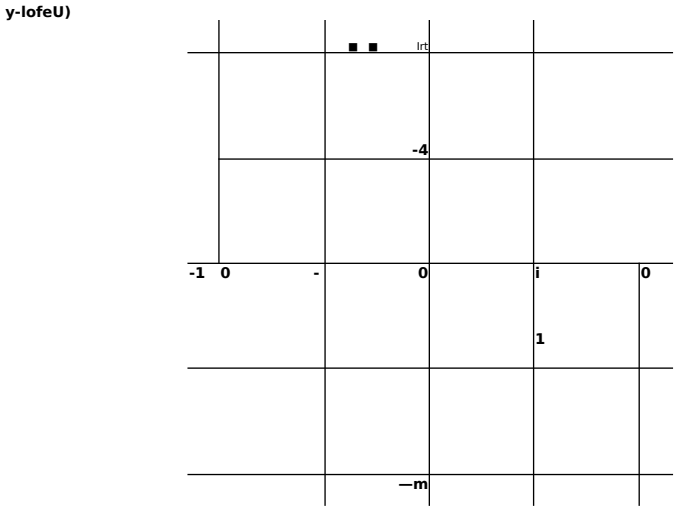
38. Prin „invers” nu vreau să spun că stările interne ale sistemului sunt ascunse de stările externe (nu-sistem), ci că *punctul de vedere* al sistemului este ascuns; e disponibil doar pentru sine. Când e vorba de problema altor minţi, un sistem nu poate cunoaşte niciodată stările interne ale altui sistem nu numai din cauză că toate stările externe sunt ascunse de el, ci şi din cauză că stările interne ale altor sisteme sunt interne exclusiv pentru acele sisteme.
39. Friston (2013), p. 8.
40. Termenul „autopoiează” a fost introdus de Maturana şi Varela (1972) pentru a defini chimia autonomă a celulelor vii.
41. Având în vedere ce am spus mai sus despre capacitatea de reprezentare a sistemului, cititorii amatori de filosofie vor observa că determină şi „intenţionalitatea” în sensul lui Brentano (1874).
42. Friston (2013), p. 2.
43. Sau, pur şi simplu, este ascunsă şi indisponibilă pentru lucru mecanic util.
44. Este ceea ce Helmholtz a numit energie care nu este liberă *(TS* în ecuaţie). Distincţia dintre energia „liberă” şi cea „legată” cu siguranţă va fi apreciată de specialiştii în opera lui Freud. Vezi capitolul 10, nota 16, despre noţiunea de „proces secundar” în

accepţiunea lui Freud.

45. $A = U + pV - TS$, unde p este presiunea şi V volumul. Această ecuaţie cuantifică energia liberă a sistemelor a căror activitate este asociată cu dilatarea şi compresia sistemului la o temperatură şi o presiune constante.
46. Ecuaţia are o formă lungă şi una scurtă. lat-o pe cea lungă: $F(s, p) = Eq[-\log p(s, 0 \mid m)] - H[q(0 \mid p)]$, unde F (energie liberă „variaţională” sau, pe scurt, energie liberă Friston) este izomorfa energiei libere Gibbs şi a energiei libere Helmholtz. În ce priveşte celelalte cantităţi din ecuaţie, s denotă stările senzoriale (ale zonei Markov, discutată mai sus), p stările interne, Eq energia medie, $p(s, 0 \mid m)$ o densitate de probabilitate cu privire la stările senzoriale şi externe (ascunse) într-un model generativ m, H denotă entropia, iar $q(0 \mid p)$ o densitate variaţională cu privire la stările ascunse parametrizate de stări interne. Relaţia dintre această ecuaţie folosită în ştiinţa informaţiei şi cea folosită în termodinamică nu este imediat evidentă. Totuşi, când ecuaţia lungă este comprimată după cum urmează, ea seamănă cu cea din termodinamică: $F = Eq - H$. Aici, F denotă energia liberă Friston, Eq energia medie, iar H entropia.
47. Adică informaţia medie obţinută după multe măsurători ale microstărilor.
48. Practic, expresiile ulterioare citate mai sus (referitoare la surprindere şi divergenţă) ne spun că, deşi energia liberă Helmholtz arată energia disponibilă pentru acţiune efectivă, energia liberă

Friston arată *diferenţa dintre modul în care lumea* respectivă *este modelată de un sistem şi modul în care lumea se comportă de fapt.* (Voi explica imediat ce legătură are acest lucru cu acţiunea.)

49. Complexitatea termenului „energie liberă Friston” are acelaşi punct fix ca energia liberă Helmholtz (presupunându-se că sistemul este închis din punct de vedere termodinamic, dar nu izolat). Dacă perturbaţiile senzoriale sunt suspendate pe o perioadă corespunzător de lungă, complexitatea este minimizată deoarece acurateţea poate fi neglijată. În acest punct, sistemul este în echilibru, iar stările interne minimizează energia liberă Helmholtz conform principiului energiei minime (care este, practic, o reformulare a legii a doua).
50. Enunţ formal: surprinderea este logaritmul negativ al probabilităţii rezultatului (*s*), într-un model dat al stărilor ascunse ale lumii ([-log *p*(*s*, *0* | *m*)] în ecuaţia de mai sus). Pe măsură ce probabilitatea (axa *x*) se îndreaptă spre 0, surprinderea (axa *y)* urcă; pe măsură ce probabilitatea se îndreaptă spre 10, surprinderea coboară.



51. Astfel, minimizarea surprinderii despre lucrurile care s-au întâmplat va minimiza, în medie, entropia. (Surprinderea este un atribut al datelor sau al observaţiilor, iar energia liberă este un atribut al credinţelor. Prin urmare, partea entropică din energia liberă nu este surprinderea medie a observaţiilor, ci entropia credinţelor despre cauzele latente ale observaţiilor.)
52. Energia liberă Friston (energia medie minus entropia) este echivalentă cu surprinderea (exprimată ca -log *p(s | m))* plus „divergenţa perceptivă” (exprimată ca *DKL* [*q*(*0 | p)* || *p*(*0 | s, m*)]), care este întotdeauna mai mare decât sau egală numai cu surprinderea. În esenţă, „surprinderea” medie înseamnă entropie (informaţională), după cum explic mai jos. „Divergenţa perceptivă” măsoară diferenţa dintre evenimentele presupuse şi evenimentele reale, într-un model generativ. *DKL* denotă divergenţa perceptivă. Vine de la divergenţa Kullback-Leibler, numită şi entropie relativă, care cuantifică divergenţa dintre două densităţi de probabilitate: cele două care ne interesează sunt despre stările ascunse – densitatea variaţională codificată în stările interne (de exemplu, neuronale) şi densitatea condiţională reală, date fiind stări senzoriale. *DKL* este întotdeauna mai mare decât sau egală cu zero. Intuim că asta se datorează faptului că logaritmii negativi (explicaţi mai jos) au întotdeauna reprezentări grafice aproximativ în formă de U, aşa că o linie ce uneşte două puncte pe U nu poate fi niciodată sub partea inferioară a U-ului (termenul

tehnic este „funcţie convexă”). Astfel se asigură că energia liberă impune surprinderii o limită în partea superioară.

53. Vezi Clark (2017): „Autovalidarea [...] apare atunci când o ipoteză explică cel mai bine unele dovezi şi, în virtutea acestui succes explicativ, aduce astfel dovezi în sprijinul propriului adevăr sau al propriei corectitudini. În asemenea cazuri, apariţia dovezii este explicată cel mai bine de ipoteză, dar faptul că dovada există este folosit pentru a veni chiar în sprijinul ipotezei”.
54. Mai este aici o particularitate tehnică, dezvăluind relaţia profundă dintre autoorganizare, teoria informaţiei şi inferenţa existenţială. Surprinderea nu este decât verosimilitate cu logaritm negativ a unor stări senzoriale, dat fiind un model în cadrul unei zone Markov. În statistica bayesiană, acest lucru e cunoscut şi sub numele (logaritm de) verosimilitate marginală *(model evidence).* Asta permite afirmaţia că minimizarea energiei libere Friston *se autovalidează,* adică maximizarea dovezilor pentru un model care generează stările senzoriale ale oricărui sistem ce are o zonă Markov.
55. De aceea, ar putea fi mai corect să spunem: „Răspunsul la această întrebare trebuie să determine comportamentul mediu al sistemului”.
56. Observaţie: acest mecanism impune constrângeri probabiliste „liberului-arbitru”. De exemplu, eşti liber să intri în cuşca leului, dar e *improbabil* că vei intra şi, dacă intri, *probabil* vei muri. Vezi, mai jos, teorema lui Bayes.

57. Unde *A* şi *B* sunt evenimente, *P(A | B)* este verosimilitatea ca *A* să apară dat fiind că *B* este adevărat (probabilitatea condiţională), *P(B | A*) este verosimilitatea ca *B* să apară dat fiind că *A* este adevărat (altă probabilitate condiţională), iar *P*(*A*) şi *P*(*B*) sunt probabilităţile de a observa *A* şi *B* independent unul de celălalt (probabilitatea marginală). Această formulare a teoremei e conformă cu Joyce (2008), care a scris *P*(*A*) | *P*(*B*) ca raport separat.
58. Teorema lui Bayes poate fi reformulată în termenii energiei libere – prin care energia liberă este descompusă în „acurateţe” şi „complexitate”. Verosimilitatea marginală este diferenţa dintre acurateţe şi complexitate, întrucât modelele cu energie liberă minimă furnizează explicaţii precise ale datelor în contextul unor costuri ale complexităţii, ceea ce, apoi, înseamnă că reducerea complexităţii modelului îmbunătăţeşte „generalizabilitatea” modelului, dar cu costul acurateţei. (În termeni bayesieni, „verosimilitatea” trebuie evaluată în raport cu „probabilitatea” pentru a preveni supraajustarea.) Knill şi Pouget (2004), p. 713, ajung la esenţa problemei: „Adevăratul test al ipotezei codificării bayesiene constă în a vedea dacă estimările neuronale ce duc la evaluări perceptive sau la comportament motor ţin seama de incertitudinea existentă în fiecare etapă a prelucrării”.
59. Pentru o descriere detaliată a microanatomiei funcţionale a acestui proces, vezi Friston (2005), Adams, Shipp şi Friston (2013) şi Parr şi Friston

(2018).

60. Astfel, de exemplu, erorile de predicţie par să fie comunicate folosind frecvenţe (înalte) din banda gama, iar predicţiile par să fie transmise prin frecvenţe (joase) din banda beta. Vezi Bastos *et al.* (2012, 2015).
61. Vezi Tozzi, Zare şi Benasich (2016): „Minimizarea energiei libere variaţionale presupune cu necesitate o codificare eficientă sub aspect metabolic corespunzătoare principiilor redundanţei minime şi transferului informaţional maxim (Picard şi Friston, 2014). Maximizarea informaţiei reciproce şi minimizarea costurilor metabolice sunt două feţe ale aceleiaşi medalii: prin descompunerea energiei libere variaţionale în acurateţe şi complexitate, putem deduce principiul informaţiei reciproce maxime ca un caz special al maximizării acurateţei, iar minimizarea complexităţii se traduce prin minimizarea costurilor metabolice (Friston *et al.*, 2015). Astfel, forma de bază a principiului energiei libere al lui Friston susţine ideea că nivelurile energetice ale activităţii cerebrale spontane, care sunt mai mici în comparaţie cu activitatea evocată, permit SNC să aibă două realizări aparent contradictorii: să minimizeze cât mai mult posibil costurile metabolice şi să maximizeze cât de mult se poate informaţia reciprocă”.
62. Vezi Clark (2015), p. 268. Totuşi, acest lucru trebuie comparat cu ceea ce spune el despre activitatea implicită a sistemului CĂUTĂRII (p. 263): „Asemenea creaturi sunt construite să-şi caute parteneri, să evite foamea şi setea şi să se implice (chiar dacă nu le este

foame sau sete) în tipurile de explorare sporadică a mediului înconjurător care le ajută să se pregătească pentru schimbări ambientale neprevăzute, penurii de resurse, concurenţi noi şi aşa mai departe. Aşadar, în orice moment, eroarea de predicţie este minimizată numai în contextul acestui set complex de «preconizări» ce definesc creatura”.

63. Friston şi Stephan (2007), p. 427.
64. M-am inspirat pentru această fabulă din metafora lui Clark (2015) despre un baraj cu fisuri.

Capitolul 8. O ierarhie predictivă

1. Hohwy (2013), p. 63.
2. Spun „centrul”, şi nu „vârful” ierarhiei deoarece pare ciudat (din punct de vedere anatomic) să plasezi nucleii hipotalamici şi ai trunchiului cerebral care monitorizează corpul *deasupra* neocortexului. Prefer să vizualizez ierarhia ca desfăşurându-se concentric, dinspre interior spre exterior, cumva ca foile unei cepe (vezi Mesulam, 2000). Până la urmă, după cum ştie orice embriolog, sistemul nervos este un tub. Să nu uităm nici că tubul neural se formează din ectoderm, prin invaginarea plăcii neurale, prin care canalul central ia locul lumii exterioare.
3. Aceasta nu este perspectiva standard asupra codificării predictive, dar tocmai de asta am şi scris cartea de faţă: arăt cum anume perspectiva noastră asupra creierului este îngreunată de cortico-centrism. *Cf.* Pezzulo (2014), p. 910: „O direcţie

interesantă pentru cercetările viitoare ar fi analizarea reglării homeostatice în cadrul inferenţei active”.

4. Clark (2015), p. 21, oferă un scurt rezumat, folosind vederea ca model: „Un anumit tipar de stimulare retiniană, întâlnit într-un context dat, ar putea fi descris cel mai bine folosind un model generativ care [...] combină reprezentări de nivel superior ale unor agenţi, obiecte, motive şi mecanisme care interacţionează, cu multe straturi intermediare ce surprind modul în care culorile, formele, texturile şi marginile se combină şi evoluează în timp. Când combinaţia unor asemenea cauze ascunse (ce se întind pe multe scări spaţiale şi temporale) se stabilizează într-un tot coerent, sistemul autogenerează datele senzoriale folosind cunoaşterea stocată şi percepe o scenă structurată ce are semnificaţie. Merită să spunem din nou că această perspectivă asupra scenei distale structurate trebuie să fie generată folosindu-se numai informaţiile disponibile din perspectiva animalului. Adică ea trebuie să fie o perspectivă înrădăcinată în totalitate în combinarea dintre orice prestructurare (a creierului şi corpului) care ar putea fi prezentă datorită istoriei evoluţiei animalului şi influenţele stimulării energetice consemnate de receptorii senzoriali. Un mijloc sistematic de a obţine asemenea perspectivă este furnizat de încercarea continuă de a autogenera semnalul senzorial folosind o arhitectură multinivel. Concret, asta înseamnă că unele conexiuni de sus în jos şi laterale dintr-un sistem

multinivel ajung să codifice un model probabilistic de cauze ce interacţionează şi funcţionează la mai multe scări în spaţiu şi timp. Recunoaştem obiecte, stări şi situaţii [...] prin identificarea setului de factori ce interacţionează (cauze distale) în cazul cărora există cea mai mare probabilitate ca combinarea lor să genereze (de aici, prezice şi explică cel mai bine) datele senzoriale de intrare”.

5. *Cf.* Gregory (1980).
6. Acest paragraf îl parafrazează pe Hohwy (2013), pp. 81-82, dar îi corectează biasul exteroceptiv.
7. Nu vreau să spun că punctele de stabilizare homeostatică nu se schimbă *niciodată,* ci doar că gradele lor de libertate sunt extrem de limitate; prin urmare, aproape întreaga alostază implică schimbare comportamentală, şi nu actualizarea elementelor anterioare înnăscute.
8. Trec aici peste chestiunea unui „Bayes empiric”. E posibil să învăţăm de la zero, dar e îngrozitor de costisitor din punct de vedere biologic.
9. Ca să ilustrez dimensiunile problemei, să reflectăm la cuvintele lui Hohwy (2013), p. 206: „Deşi ne spune întreaga poveste a erorii de predicţie, nu pare să existe nicio contradicţie în ideea că există o creatură dotată cu tot aparatul necesar pentru minimizarea erorii de predicţie, care se implică în aceeaşi inferenţă predictivă ca şi noi – indiferent de gradul de detaliere naturală, neuronală pe care vrem să-l specificăm –, dar care nu este conştientă din punct de vedere fenomenologic. Desigur, ne-am aştepta ca

această creatură să fie conştientă, dar nimic din toată povestea ei fizică nu impune să şi fie conştientă. Asta înseamnă că rămâne deschisă întrebarea dacă conştiinţa este ceva dincolo de ce este fizic sau nu”.

Cum ar putea să *nu* fie conştientă o creatură având exact acelaşi aparat neuronal ca tine şi ca mine? Îndrăznesc să afirm că „zombiului filosofic” imaginat aici de Hohwy îi lipseşte aparatul pentru sentimente. Pentru o interpretare alternativă a problemei pe care o pune el, vezi capitolul 12.

Capitolul 9. De ce şi cum se naşte conştiinţa

1. Rolls (2019), p. 10, sublinierea mea. Vezi şi comentariul meu:
   Solms (2019b).
2. „Nu trebuie folosite mai multe lucruri decât e necesar.” Dacă
   există mai multe moduri posibile de a explica ceva, cel care foloseşte cele mai puţine presupuneri este probabil cel corect.
3. Roger Sessions, *New York Times,* 8 ianuarie 1950.
4. Friston (2009), p. 299, defineşte rolul preciziei ca fiind de „a controla influenţa relativă a aşteptărilor anterioare la diferite niveluri”. După cum arată Hohwy (2013), p. 199: „Dacă se creşte amplificarea [precizia] unui semnal, atunci trebuie redusă amplificarea altor semnale. Altfel, ideea de amplificare îşi pierde semnificaţia: suma ponderilor trebuie să fie unu. Aşadar, pe măsură ce aşteptările

pentru precizie cresc amplificarea unei erori de predicţie, amplificarea altora se va reduce”. *Cf.* Clark (2016), p. 313: „Feldman şi Friston (2010) arată că precizia se comportă ca şi cum ar fi ea însăşi o resursă limitată, în sensul că creşterea preciziei pentru unele erori de predicţie necesită reducerea ei pentru altele. De asemenea, e interesant comentariul autorilor *(op. cit.,* p. 11): «Motivul pentru care precizia se comportă ca o resursă este că modelul generativ conţine credinţe anterioare conform cărora precizia logaritmică este redistribuită prin canale senzoriale într-o modalitate sensibilă la context, dar este conservată în toate canalele»”. Aceste afirmaţii amintesc de cele speculate de mine în capitolul 6 despre fiecare neuromodulator ce ajustează diferite aspecte ale activării. E interesant că oscilaţiile (erorile) gama răspund la acetilcolină.

5. Neurofiziologii numesc precizia şi „amplificare”. Terminologia foarte variată folosită de specialişti din diverse discipline poate fi derutantă aici. „Activitatea” sinaptică (= neurotransmisia) este modulată de „amplificarea” postsinaptică (= neuromodularea), care, în timp, determină „eficienţa” (= neuroplasticitatea). Aceleaşi variabile sunt descrise de specialiştii în ştiinţe computaţionale drept „stări” de semnal, „precizii” de semnal şi, respectiv, „parametri” de semnal. În general, neurotransmisia necesită stări, neuromodularea necesită precizii şi plasticitatea necesită parametri.
6. Friston numeşte acest lucru „căutare epistemică a hranei”.

7. Acest paragraf îl parafrazează îndeaproape pe Clark (2015), p. 70.
8. De aici apare „orbirea perceptivă”, cel mai cunoscut exemplu găsindu-se pe https://www.youtube.com/watch?v=vJG698U2Mvo (vezi Chabris şi Simons, 2010).
9. Varela, Thompson şi Rosch (1991), p. 198. Varela numeşte „institutivă” această abordare a percepţiei. Vezi Clark (2015), p. 173: „Preocuparea generală a unei abordări institutive a percepţiei nu este să determine cum urmează să fie recuperată o lume independentă de observator, ci să determine principiile comune sau legăturile legitime dintre sistemele senzoriale şi motoare ce explică modul în care acţiunea poate fi ghidată perceptiv într-o lume dependentă de observator”.
10. Există un volum mare de literatură experimentală care susţine empiric această conceptualizare formală a preciziei. Deocamdată voi menţiona un singur studiu, care ilustrează ideea de bază. Folosind RMNf, Hesselmann şi colaboratorii săi (2010) au analizat activitatea de stare iniţială din creier în condiţii care manipulau aşteptările legate de precizie. Autorii le-au prezentat participanţilor la cercetare stimuli (vizuali şi auditivi) cu diferite niveluri de zgomot; când participanţii se aşteptau la precizie, a fost crescută amplitudinea, iar când se aşteptau la imprecizie, amplitudinea a fost redusă. Adică predicţia de sus în jos a fost facilitată atunci când erau preconizate semnale senzoriale imprecise. Vezi Feldman şi Friston (2010) pentru o descriere

completă a modului în care optimizarea preciziei explică atât atenţia endogenă, cât şi atenţia exogenă.

11. După cum îţi aminteşti de la p. 45, Freud a scris următoarele la vremea respectivă, într-o scrisoare către Fliess, din 20 octombrie 1895: „Într-o noapte plină [...] barierele au fost brusc ridicate, vălurile au căzut şi am putut să înţeleg de la detaliile nevrozelor la determinanţii conştiinţei. Totul părea să se potrivească, rotiţele erau angrenate, obiectul îţi dădea impresia că era într-adevăr o maşină care, curând, va merge de la sine”. Şi a continuat: „Cele trei sisteme de neuroni [*0*, *0* şi *to*], condiţiile de cantitate libere şi fixe [*Q^*], procesele primare şi secundare, tendinţa principală şi tendinţa de compromis a sistemului nervos, cele două reguli biologice ale atenţiei şi apărării, indicatorii calităţii [*to*], realităţii şi gândirii, starea grupurilor psihosexuale, determinarea sexuală a represiei şi, în final, determinanţii conştiinţei ca funcţie perceptivă – toate acestea se potriveau şi încă se potrivesc! Desigur, nu îmi pot ascunde încântarea”. Totuşi, la scurt timp, a realizat că rotiţele nu erau angrenate, lucrurile nu se potriveau complet. Atunci a abandonat proiectul. Când eu şi Friston l-am reînviat – şi, luând în calcul mai bine de un secol de progrese în neuroştiinţe, sperăm că l-am finalizat (cel puţin în mare; vezi Solms, 2020b) –, mi-am amintit de remarca dură a lui Freud: „Obiectul îţi dădea impresia că era într-adevăr o maşină care, curând, va merge de la sine”. Vezi capitolul 12.
12. De observat că *w* nu este un vector, ci un scalar

(măsoară o matrice a preciziei). De observat şi că termenii Q and Qq au fost utilizaţi uşor diferit în Solms şi Friston (2018) faţă de cum îi folosesc eu aici.

13. Unde e este un vector, iar F un scalar. Notaţia cu puncte în ecuaţiile de mai jos implică un produs scalar (adică înmulţire a matricii sau a vectorului). Aceste ecuaţii reprezintă în întregime contribuţia lui Friston la colaborarea noastră.
14. $F = -\log P(0(M))$.
15. $E[F] \sim E[-\log P(0)] = H[P(0)] = -¥2 \bullet \log(|w|)$, unde $E[-]$ denotă preconizarea sau media, iar P şi H probabilitatea şi, respectiv, entropia (ca şi înainte); în baza premiselor gaussiene despre fluctuaţia aleatorie.
16. Importanţa generală a optimizării preciziei a fost recunoscută cu mult timp în urmă, mai ales în legătură cu atenţia (vezi Friston, 2009; Feldman şi Friston, 2010). Totuşi, primul specialist care i-a recunoscut importanţa pentru conştiinţă a fost Katerina Fotopoulou (2013), p. 35: „Un aspect esenţial al conştiinţei poate servi la consemnarea proprietăţii anterior menţionate a «incertitudinii» şi a proprietăţii sale inverse, precizia. Această opinie contrazice vechea perspectivă intuitivă asupra conştiinţei afective esenţiale ca proprietate hedonică de monitorizare, exprimată de Solms în termeni freudieni ca seria plăcere- neplăcere. În schimb, eu afirm că proprietatea esenţială a

acestui aspect al conştiinţei (ca fiind opus conştiinţei perceptive [...]) este un fel de principiu al certitudinii-incertitudinii sau al dezambiguizării". Ţin să precizez că, în ce priveşte plăcerea-neplăcerea versus certitudinea- incertitudinea, cred că Fotopoulou m-a înţeles puţin greşit, de unde şi răspunsul meu: „Nu sunt sigur că [...] înţeleg la ce se referă când afirmă că afectul monitorizează incertitudinea, şi nu proprietatea hedonică. În opinia mea, proprietatea hedonică *este* măsura incertitudinii" (Solms, 2013, p. 81).

17

$$\frac{\partial}{\partial t}M = \frac{\partial F}{\partial SM} \qquad \frac{\partial F}{\partial SM} = \frac{\partial e}{\partial SM}\,k\,e \quad (1)$$

$$\frac{d}{dt}Qn = \frac{\partial F}{\partial SQT} \qquad \frac{\partial F}{\partial SQT} = \frac{\partial e}{\partial SQT}\,\Pi\,\frac{\partial e}{\partial SQT} \quad (2)$$

$$\frac{d}{dt}\Pi = \frac{\partial F}{\partial \mathrm{Sro}} = \frac{1}{2}\cdot(\omega^{-1} - e\cdot e) \quad (3$$

Unde *d* denotă un derivat parţial şi *t* timpul, iar eroarea de predicţie şi energia liberă sunt:

$$e = \langle p(M) - i|j(QT)\rangle$$

$$F = \frac{1}{2}(\mathrm{era}\text{-}e\text{-}\log(<o))$$

Prezint în mare aceste ecuaţii, ca nişte desene cu pretenţii. Ele trebuie să fie dezvoltate suplimentar, iar această elaborare se poate face numai atunci când sunt implementate (vezi capitolul 12). De exemplu, după cum am afirmat mai sus, în abordări mai complexe, ar trebui să avem în vedere şi modelele generative ierarhice (cu precizii la fiecare nivel) şi să includem incertitudinea condiţională despre stările externe. Mai mult, ecuaţiile pun laolaltă toate erorile senzoriale de predicţie –

incluzând modalităţile exteroceptive, proprioceptive şi interoceptive. De observat că, aici, termenul „proprioceptiv” este sinonim cu „chinestezic” (Friston foloseşte „proprioceptiv” pur şi simplu pentru armonia aliterativă cu „exteroceptiv” şi „interoceptiv”).

18. În termeni tehnici, acest lucru se numeşte „algoritmul gradientului descendent”, unde gradientul este rata de schimbare a energiei libere cu precizie.
19. În baza asumpţiilor noastre simplificatoare despre codificarea credinţelor bayesiene (vezi mai sus).
20. În acest sens, se poate spune că precizia joacă rolul „demonului lui Maxwell” – un experiment mental creat de fizicianul James Clerk Maxwell (1872): un demon controlează o uşă mică între două încăperi cu gaze. Moleculele de gaz plutesc cu viteze diferite. Când cele mai rapide din prima încăpere ajung la uşă, demonul o deschide şi o închide, pentru foarte scurt timp, astfel încât ele trec în a doua încăpere, iar cele mai lente rămân în urmă. Întrucât moleculele mai rapide generează mai multă căldură decât cele mai lente, acest lucru va scădea entropia – ceva ce nu se poate întâmpla fără lucru mecanic. Dacă punem semnul egalităţii între trecerea moleculelor din această analogie şi neurotransmisia semnalelor senzoriale, atunci cântărirea preciziei (neuromodularea) face ce face demonul lui Maxwell: selectează semnale senzoriale pentru a încurca a doua lege a termodinamicii. De observat că, în termenii dinamicii sistemului descrise aici, conştiinţa

este *activitatea demonului lui Maxwell,* şi nu *trecerea moleculelor* facilitată de activitate. Asta înseamnă că conştiinţa este *optimizarea preciziei în raport cu energia liberă,* şi nu *trecerea mesajelor printr-o ierarhie predictivă.* Prin urmare, în figura 17 conştiinţa este activitatea *w* (precizia), care determină influenţa relativă a *e* (semnalele de eroare) asupra *Qq* (modelul intern). Demonul-precizie modelează influenţa erorilor în legătură cu modelul. Conştiinţa depinde de un atribut al credinţelor, şi nu de conţinutul credinţelor (adică depinde de precizia fluctuantă a credinţelor despre situaţiile interne şi externe [sau de încrederea în ele]). Activitatea acestui demon *cauzează* sechele senzoriale prin amplificarea sau atenuarea erorilor de predicţie; optimizarea preciziei nu este inerentă semnalelor de eroare în sine. Pentru mai multe despre implicaţiile biologice ale demonului lui Maxwell, vezi excelenta carte a lui Paul Davies (2019) care, din păcate, a apărut după ce am realizat noi aplicarea conceptului la conştiinţă (Solms şi Friston, 2018), aşa că nu am citat-o.

21. Brown *et al.* (2013), Feldman şi Friston (2010), Frith, Blakemore şi Wolpert (2000).
22. Cisek şi Kalaska (2010), Frank (2005), Friston *et al.* (2012), Friston, Schwartenbeck, FitzGerald *et al.* (2014), Moustafa, Sherman şi Frank (2008).
23. Hohwy (2013), Seth (2013), Ainley *et al.* (2016). Pentru sensibilitatea interoceptivă şi modularea socială a durerii, vezi şi Crucianelli *et al.* (2017), Fotopoulou şi Tsakiris (2017), Krahe *et al.* (2013),

Decety şi Fotopoulou (2015), Paloyelis *et al.* (2016) şi Von Mohr şi Fotopoulou (2017).

24. Ferrarelli şi Tononi (2011), Lisman şi Buzsaki (2008), Uhlhaas şi Singer (2010).
25. Vezi Hobson (2009) şi Hobson şi Friston (2012, 2014). Totuşi, în opinia mea, o explicaţie satisfăcătoare a viselor în acest cadru trebuie să înceapă de la caracterul lor *conştient* (şi afectiv).
26. Nour şi Carhart-Harris (2017).
27. Dehaene şi Changeux (2011), Friston, Breakspear şi Deco (2012).
28. Montague *et al.* (2012), Corlett şi Fletcher (2014), Friston, Stephan, Montague şi Dolan (2014), Wang şi Krystal (2014).
29. Pentru o discuţie despre „empatie”, vezi Solms (în tipar). Termenul german este *Einfühlung,* care literal înseamnă „simţire în”.

Capitolul 10. Înapoi la cortex

1. Zeki (1993), p. 236, sublinierea mea.
2. *Ibidem,* p. 238.
3. Nu intenţionez să implic că acesta este un proces cognitiv de nivel superior. Cortexul vizual „de proiecţie” pare să umple pata oarbă cu orice o înconjoară. În condiţii naturale, la asta contribuie mişcările frecvente ale ochilor, care se asigură că ai aproape mereu o imagine reală (memorie pe termen ultrascurt) a ceea ce este în pata oarbă. Pentru diversitatea mecanismelor implicate, vezi

Ramachandran (1992), Ramachandran şi Gregory (1991) şi Ramachandran, Gregory şi Aiken (1993). Se întâmplă că, pe lângă umplerea cu conţinut a petei oarbe, are loc şi eliminarea obiectelor nedorite, cum ar fi umbrele „musculiţelor” entoptice şi vascularizaţia retiniană.

4. Suprapunerea dintre câmpurile reprezentate de fiecare ochi anulează petele oarbe, ceea ce înseamnă că trebuie să închizi un ochi ca să observi umplerea iluzorie menţionată.

5. Există, desigur, o anumită suprapunere între ele, dar măsura în care obiectele atenţiei vizuale sunt literalmente secţionate în două este uşor de demonstrat în cazurile de neglijare unilaterală (stânga) în urma unor leziuni la nivelul emisferei drepte.

6. Întâmplător, aceasta este o dovadă pentru opinia că funcţia „de legare” a conştiinţei vine de sub cortex, din trunchiul cerebral unitar, şi nu din cele două emisfere cerebrale (vezi Panksepp şi Biven, 2012).

7. Solms *et al.* (1988).

8. L-a descris drept „un teatru minunat” (Helmholtz, 1867, p. 776).

9. *Ibidem,* p. 438.

10. Majoritatea oamenilor care lucrează în paradigma codificării predictive nu acordă suficientă atenţie faptului că inferenţa perceptivă este un proces *inconştient.* De exemplu, eu nu sunt de acord cu Hohwy (2013) când spune că „suntem conştienţi de «fantezia» generată de modul în care predicţiile curente atenuează eroarea de predicţie”. În opinia mea, nu *devenim conştienţi* de „fanteziile” predictive

dacă nu *se ciocnesc* cu realitatea. Şi slavă Domnului pentru acest lucru.

Se pare că Clark este de acord cu mine în această privinţă (Lupyan şi Clark, 2015, p. 281, sublinierea mea): „Deşi majoritatea predicţiilor sunt inconştiente, uneori le putem conştientiza *când sunt încălcate.* De exemplu, imaginează-ţi că bei dintr-un pahar în care crezi că este suc de portocale, dar realizezi, după ce îl guşti, că, de fapt, e lapte. Diferenţa dintre gustul de lapte când te aştepţi să fie lapte şi când, în loc de lapte, te aştepţi să fie suc de portocale este conştientizarea aşteptării referitoare la sucul de portocale

(Lupyan, 2015). În mod similar, gândeşte-te la experienţa unei omisiuni nepreconizate, ca atunci când lipseşte o notă muzicală dintr-o compoziţie familiară. Asemenea omisiuni pot fi la fel de frapante sub aspect perceptiv şi la fel de predominante ca tonul cel mai vibrant – altfel, un efect uimitor ce se explică clar dacă presupunem că construcţia experienţei perceptive implică aşteptări bazate pe vreun fel de model a ceea ce e probabil să se întâmple”. În pofida afirmaţiilor lui Hohwy (2013), uneori el se apropie de opinia mea. Într-un articol anterior, a făcut următoarea afirmaţie (Hohwy, 2012, p. 11, sublinierea mea): „Această semnătură temporală corespunde codificării predictive în măsura în care, când eroarea de predicţie a unui stimul este suprimată complet şi nu mai are loc nicio explorare (întrucât inferenţa activă este atenuată, din cauză că sistemul e ocupat temporar pe parcursul

atenţiei defocalizate), probabilitatea ar trebui să înceapă să scadă. Ajungem aici pornind de la ideea că *ceea ce impulsionează percepţia conştientă este procesul concret al suprimării erorii de predicţie”.* Totuşi, în cartea lui (2013, p. 201, sublinierea mea), spune: „Percepţia conştientă este efectul inferenţei perceptive inconştiente. Nu ne implicăm în mod conştient într-o actualizare bayesiană a predicţiilor noastre anterioare în lumina unor dovezi noi şi nici a modului în care inputul senzorial este prezis şi apoi atenuat. Ceea ce este conştient e rezultatul inferenţei – *concluzia”.* De aceea nu sunt de acord cu Hohwy (2013), dacă înţeleg bine ce vrea să spună. Dar suntem amândoi de acord că *travaliul* constând în minimizarea semnalelor de eroare precise generează conştiinţa, adică discrepanţa problematică dintre predicţie şi eroarea cântărită de precizie. Dar nu sunt de acord cu el când spune că „suntem conştienţi de «fantezie»”. În cel mai bun caz, Hohwy pare să creadă că suntem conştienţi de „fantezia” ce încearcă *să suprime* eroarea de intrare, dar eu cred că suntem conştienţi de faptul că *nu* suprimă eroarea – adică suntem conştienţi de „lucrul în desfăşurare” predictiv cauzat de discrepanţă. Aşa devine realitatea predominantă. Poate că această idee este doar o chestiune de semantică. Pentru mine, contează că conştiinţa în percepţie este impulsionată de *incertitudine,* şi nu de *cele-mai-bune-presupuneri* generate de certitudinea relativă. După cum văd eu situaţia, presupunerile devin conştiente numai atunci când sunt nesigure şi se retrag când sunt confirmate.

Cu alte cuvinte, devenim conştienţi de fanteziile noastre numai atunci când sunt contrazise de realitate. Conştiinţa ar putea fi descrisă ca un proces de *dezambiguizare.* Sunt conştient că această chestiune este complicată de diferenţa dintre inferenţa perceptivă şi cea activă (şi dintre atenţia exogenă şi cea endogenă). Aşa se explică de ce există o literatură bogată – cea pe care se bazează Hohwy (2013) – care sugerează că percepem în mod preferenţial (conştient) ceea ce se potriveşte cu aşteptările noastre *şi* există şi alte lucrări de specialitate care sugerează că percepem în mod preferenţial (conştient) ceea ce este cel mai nepreconizat. Pentru mine, această contradicţie este rezolvată de conceptul *prioritizării afectului,* care este impulsionată, în ambele cazuri, de echilibrul dintre ceea ce eu numesc în sens larg „nevoi” şi „oportunităţi” (vezi capitolul 5). Pe scurt, percepem conştient ceea ce este predominant *în relaţie cu nevoia prioritizată pe moment.* Nevoile (afectele) prioritizate produc, prin definiţie, cele mai precise semnale de eroare. Opiniile mele în această privinţă sunt elaborate în secţiunea următoare. Totuşi, în cele din urmă, aceasta este o problemă empirică. În prezent, împreună cu studenţii mei Donne van der Westhuizen şi Julianne Blignaut şi cu fostul meu student Joshua Martin, o studiez folosind paradigma rivalităţii binoculare standard, ce se referă la *conştiinţa* perceptivă. Vezi şi Pezzulo (2014), Yang, Zald şi Blake (2007) şi Stein şi Sterzer (2012).

11. Am publicat pentru prima dată această idee în Solms

(2013), apoi am revenit asupra ei în diverse moduri, de exemplu, Solms (2015b, 2017b, 2017c, 2018b).

12. Freud (1920), p. 25.

13. Freud a atribuit conştiinţă şi memorie unor sisteme diferite de neuroni (w şi, respectiv, 0) care, ulterior, au devenit sistemele lui metapsihologice Cs şi Pcs – dar a considerat că *amândouă* sunt sisteme corticale. Pasajul din care este citată fraza „conştiinţa ia naştere acolo unde se sfârşeşte urma mnezică” afirmă foarte clar acest lucru, deşi a fost scris la 25 de ani după „Proiect” (vezi Freud, 1920, pp. 25 şi urm.).

14. Bargh şi Chartrand (1999), p. 476: „Unele dintre sistemele de ghidare automată pe care le-am evidenţiat sunt «naturale» şi nu necesită experienţă pentru a se dezvolta. Acestea sunt, pe de o parte, asocierea reprezentărilor perceptive şi comportamentale şi conexiunea dintre procesele de evaluare automată şi, pe de altă parte, dispoziţia şi comportamentul. Alte forme de autoreglare automată apar din experienţa repetată şi consecventă. Ele se conectează la regularităţile experienţei individului şi preiau sarcini de la ghidarea şi alegerea conştiente atunci când alegerea respectivă nu este exercitată cu adevărat. Astfel, obiectivele şi motivele pot ajunge să funcţioneze nonconştient în anumite situaţii, stereotipurile pot fi asociate multă vreme cu caracteristicile perceptive ale unor grupuri sociale, iar evaluările pot fi integrate cu reprezentările perceptive ale persoanei, obiectului sau

evenimentului astfel încât devin active imediat şi neintenţionat în cursul percepţiei”.

15. Din punct de vedere tehnic, stările neuronale, care sunt intensificate de preciziile neuronale, actualizează parametrii neuronali.
16. Asta coincide perfect cu ideea lui Freud de „proces secundar”. El a conceput procesele secundare drept o „legare” a formei primare de energie pulsională, care este „mobilă”. Legarea energiei libere Friston reprezintă fundamentul mecanic a ceea ce numim *travaliu* psihic (lucru mecanic util). Această concluzie este foarte importantă: „Eu cred că această deosebire [dintre energia liberă şi energia latentă] prezintă până acum înţelegerea cea mai profundă pe care o avem despre natura energiei nervoase şi nu văd cum s-ar putea ca cineva s-o evite” (Freud, 1915b, p. 188) [ed. rom.: „Inconştientul”, traducere de G. Lepădatu, în *Opere esenţiale,* vol. 3: *Psihologia inconştientului.* Bucureşti: Trei, e-book]. De reţinut totuşi că Freud considera că procesul secundar poate să funcţioneze şi *în mod preconştient*. De aici apare o întrebare empirică: poate procesarea corticală să-şi îndeplinească funcţia de stabilizare în lipsa conştiinţei? Această întrebare aminteşte de controversa actuală cu privire la memoria de lucru nondeclarativă (vezi Hassin *et al.,* 2009).
17. Nu vreau să sugerez că orice reconsolidare aduce cu sine conştientizarea *memoriei aflate în curs de actualizare* – ea poate aduce cu sine modularea afectivă a procesului de actualizare ce rămâne inconştient, din punct de vedere cognitiv (vezi mai

jos). Se întâmplă că mecanismele pe care tocmai le-am menţionat explică fenomenul psihologic pe care Freud l-a numit „rezistenţă”: reticenţa noastră aparte de a ne actualiza modelele predictive în faţa unor dovezi contradictorii. (Din păcate, această afirmaţie e valabilă chiar şi în cazul unor specialişti!)

18. Termenul „recompensă = predicţie” din diagrama lui Schultz i-ar putea deruta pe unii cititori. Atunci când consecinţele senzoriale ale unei acţiuni se potrivesc cu consecinţele prezise, nu se întâmplă nimic, nu există nicio „recompensă”. Senzaţiile, atât negative, cât şi pozitive, semnifică întotdeauna erori (vezi figura 12). Totuşi, în limbaj behaviorist, „recompensă” nu implică vreo senzaţie. Înseamnă doar că predicţia este consolidată. Adică, folosind terminologia mea, îi este atribuită o precizie mai mare. Pentru mine, asta înseamnă că senzaţia va fi resimţită drept plăcută *dacă predicţia relevantă a fost prioritizată de triunghiul de decizie din trunchiul cerebral.*
19. Misanin, Miller şi Lewis (1968).
20. Nader, Schafe şi LeDoux (2000). Vezi Dudai (2000) pentru o prezentare accesibilă. Reconsolidarea este strâns legată de conceptul de „retranscriere” a memoriei în accepţiunea lui Freud. Vezi scrisoarea lui către Fliess datată 6 decembrie 1896, în care anticipează, cel puţin, conceptul de consolidare a sistemelor: „După cum ştii, lucrez la ipoteza că mecanismul nostru fizic ia naştere printr-un proces de stratificare: materialul prezent sub forma unor urme mnezice care este supus din când în când unei

rearanjări corespunzătoare unor condiţii noi – supus unei *retranscrieri.* Astfel, nouă la teoria mea este teza că memoria e prezentă nu o dată, ci de mai multe ori, că este exprimată în diferite tipuri de indicaţii. Am postulat un tip similar de rearanjare cu ceva timp în urmă *(Afazia* [Freud, 1891]) pentru căile ce pleacă din periferie. Nu pot să spun câte asemenea consemnări există: cel puţin trei, probabil mai multe. Asta se vede în următoarea imagine schematică [în care conştiinţa apare drept «consemnarea» finală a urmei; vezi figura 8)], în care se presupune că consemnările diferite sunt şi separate (nu neapărat topografic) în funcţie de neuronii ce le sunt vehicule. Poate că această ipoteză nu e necesară, dar e cea mai simplă şi poate fi acceptată provizoriu". Freud a conceptualizat „represia" ca pe un eşec de retranscriere. Vezi Solms (2017c), pentru o actualizare din perspectivă neuropsihanalitică.

21. Astfel, atât potenţarea pe termen lung, cât şi depresia pe termen lung sunt modulate de sistemul reticulat activator (vezi Bienenstock, Cooper şi Munro, 1982). De unde şi capacitatea ECT şi a crizelor epileptice, toate acţionând prin intermediul sistemului reticulat activator, de a interfera cu consolidarea memoriei.
22. De fapt, Hebb (1949), p. 62, a spus: „Să presupunem că persistenţa sau repetarea unei activităţi cu reverberaţie (sau «urmă») tinde să inducă modificări celulare de durată care contribuie la stabilitatea ei [...]. Atunci când un axon al celulei *A* este suficient de

aproape încât să excite o celulă *B* şi ia parte în mod repetat sau persistent la descărcarea ei, are loc un anumit proces de creştere sau o schimbare metabolică într-o celulă sau în amândouă, astfel încât eficienţa celulei *A,* ca una dintre celulele care o descarcă pe *B,* creşte".

23. În prezent, planific împreună cu Cristina Alberini o serie de experimente privind rolurile PAG şi ale sistemului reticulat activator în învăţarea conştientă versus învăţarea inconştientă. Aceste experimente vor clarifica rolurile diverşilor nuclei „ai activării" din trunchiul cerebral superior care modulează cogniţia conştientă din prozencefal. Paradigma reconsolidării promite să dezvăluie unele dintre mecanismele intracelulare elementare ale conştiinţei perceptive în relaţie cu învăţarea. De exemplu, atât consolidarea, cât şi reconsolidarea pot fi perturbate de inhibarea sintezei proteice şi amândouă necesită factorul CREB de transcripţie a genelor. Totuşi, cercetările recente sugerează că în nucleul amigdalian este necesar factorul neurotrofic derivat din creier (BDNF) pentru consolidare, dar nu şi pentru reconsolidare, precum şi că Zif268, factorul de transcripţie şi genă precoce imediată, este necesar pentru reconsolidare, dar nu şi pentru consolidare. O dublă disociere similară între Zif268 pentru reconsolidare şi BDNF pentru consolidare s-a descoperit în hipocamp. Vezi Debiec *et al.* (2006) şi Lee, Everitt şi Thomas (2004).
24. Riggs şi Ratliff (1951), Ditchburn şi Ginsborg (1952). E instructiv să ne amintim observaţiile lui Helmholtz

despre atenţie (1867, p. 770): „Starea naturală neforţată a atenţiei noastre este de a hoinări continuu spre alte lucruri, astfel încât, atunci când se epuizează interesul faţă de un obiect, când nu mai putem percepe nimic nou, atunci, împotriva voinţei noastre, atenţia ni se îndreaptă spre altceva [...]. Dacă vrem să ne menţinem atenţia asupra unui obiect, trebuie să tot găsim ceva nou la el, mai ales dacă alte senzaţii puternice încearcă să o deturneze”. Aceste afirmaţii trebuie puse în legătură cu am spus mai sus despre CĂUTAREA implicită.

25. Vezi https://en.wikipedia.org/wiki/Lilac_chaser
26. Dacă îţi spune cineva că aceste efecte nu implică creierul (cu alte cuvinte, că ele apar numai la nivelurile periferice ale sistemului nervos), zi-i să citească Hsieh şi Tse (2006). Vezi şi Coren şi Porac (1974).
27. Vezi Oberauer *et al.* (2013).
28. Exclud de aici „amorsajul” şi învăţarea perceptivă. Acestea necesită într-adevăr imagini şi implică cortexul. Majoritatea celorlalte afirmaţii pe care le fac despre memoria nondeclarativă nu sunt valabile în cazul amorsajului şi învăţării perceptive, care aparţin altei categorii; ele sunt „eşafodajul” memoriei declarative. Vezi ce spun mai departe despre funcţia de amorsaj a cuvintelor, de exemplu.
29. Ganglionii bazali nu sunt afectaţi la unii copiii cu hidroanencefalie (dar nu la toţi) şi la unele animale supuse procedurii de decorticare cerebrală.
30. Din punct de vedere tehnic, complexitatea este entropia relativă dintre credinţele posterioare şi cele

anterioare sau dintre distribuţiile probabilităţii în stările externe. Această definiţie a complexităţii decurge din faptul că verosimilitatea marginală este diferenţa dintre acurateţe şi complexitate (vezi capitolul 8). Întrucât verosimilitatea marginală este sporită în mod activ prin minimizarea energiei libere, acurateţea predicţiilor creşte, ca şi complexitatea lor. Cu alte cuvinte, creşterea complexităţii modelului este întotdeauna permisă de capacitatea de a face predicţii mai precise – cum se întâmplă de regulă în sistemele de memorie corticale.

31. Compară cu Hohwy (2013), p. 202: „Ideea ar fi că inferenţa perceptivă circulă într-un spaţiu determinat atât de acurateţea erorii de predicţie, cât şi de preciziile erorii de predicţie. Putem descrie acest lucru într-un mod mai simplu, dacă considerăm că acurateţea erorii de predicţie creşte odată cu amplitudinea inversă a erorii de predicţie în sine, precum şi că precizia erorii de predicţie creşte odată cu amplitudinea inversă a fluctuaţiilor aleatorii legate de incertitudinea predicţiilor [...]. Acest lucru justifică o descriere unitară a relaţiei dintre percepţia conştientă şi atenţie. Relaţia dintre ele este ca între o inferenţă statistică de gradul întâi şi una de gradul doi”. Vezi şi Hohwy (2012), discutat pe scurt mai sus.
32. Această secţiune se bazează masiv pe Clark (2015), căruia îi sunt recunoscător.
33. Diferenţa se învârte în jurul faptului că visele aproape că nu conţin semnale de eroare exteroceptive, datorită schimbărilor dramatice la

nivelul cântăririi preciziei care apar odată cu instalarea somnului (Honson şi Friston, 2012, 2014). Clark descrie acest lucru drept „izolarea faţă de divertisment”, dar vezi nota 35, mai jos.

34. Clark (2015), p. 273.
35. Totuşi, observă că creierul în starea de somn nu este „în izolare faţă de recreere”, după cum afirmă Clark (2015), p. 107. Principalul „semnal senzorial de impulsionare” al creierului este întotdeauna endogen. Pur şi simplu, acest semnal nu poate să te izoleze de divertisment – adică de nevoile biologice. Dacă noi, specialiştii în neuroştiinţe, continuăm să trecem cu vederea acest fapt fundamental, nu vom înţelege niciodată viaţa psihică şi locul ei în natură.
36. Vezi Domhoff (2017).
37. Clark (2015), p. 274.
38. Vezi Solms (2020a) pentru o dezvoltare a acestei idei.
39. Asta se întâmplă şi în „starea de repaus” a modului implicit.
40. E uşor să uităm că hipocampul face parte din sistemul limbic – creierul emoţional.
41. Okuda *et al.* (2003), Szpunar *et al.* (2007), Szpunar (2010), Addis *et al.* (2007).
42. Ingvar (1985).
43. Schacter, Addis şi Buckner (2007), p. 660, sublinierea mea.
44. Faptul că un flux *schimbător* (impredictibil) suprimă o imagine *constantă* (mai predictibilă) este, desigur, interesant în sine.
45. Lupyan şi Ward (2013), p. 14196, sublinierea mea.

În termenii aspectelor tehnice referitoare la percepţia conştientă discutate mai sus, acesta este, desigur, un exemplu de atenţie endogenă. În acest sens, vezi şi definiţia tehnică a „predominanţei” de la p. 214. Cuvintele stimulează predominanţa în mod artificial. Totuşi, amorsajul verbal este anulat cu uşurinţă de o surpriză exogenă puternică (adică precisă). Cu alte cuvinte, prioritizarea de jos în sus a nevoilor la nivelul triunghiului de decizie din mezencefal domină în mod invariabil procesele de sus în jos din prozencefal.

46. Lupyan şi Thompson-Schill (2012).
47. Qukur *et al.* (2013). Vezi din nou experimentul despre orbirea perceptivă citat în capitolul 9, nota 8.
48. Clark (2015), p. 286.
49. Roepstorff şi Frith (2004).
50. Vezi Zhou *et al.* (2017).
51. Panksepp şi Biven (2012), p. 396: „Iniţial, *experienţele* audiovizuale conştiente au fost în mare parte afective (Panksepp, 1998). Rapiditatea cu care stimulii vizuali şi auditivi subiţi pot să ne surprindă şi să ne sperie, mai ales când asemenea stimuli îşi au originea foarte aproape de corpul nostru, sugerează o integrare primară profundă a acestor sisteme senzoriale cu unele dintre mecanismele noastre afective de supravieţuire esenţiale. Gândeşte-te şi că suntem înclinaţi să asociem anumite culori cu sentimente”.
52. Gândeşte-te şi la conceptul tot mai influent al „atingerii afective”.
53. Nu spun că conştiinţa perceptivă *în totalitatea ei* nu

are forţă cauzală. Îşi dobândeşte forţa tocmai deoarece
contextualizează afectul.

54. Vezi Hurley, Dennett şi Adams (2011), o carte despre umor asupra căreia Dennett mi-a atras atenţia recent, deoarece ajunge la concluzii despre arhitectura funcţională a minţii remarcabil de similare cu ale mele.
55. Lungimile de undă şi intensităţile auditive şi vizuale sunt măsurate, comparate şi clasificate constant de cortex, atât conştient, cât şi inconştient.
56. *Cf.* Clark (2015), p. 207: paradigma energiei libere sugerează „nu că trăim ca atare propriile noastre semnale de eroare în ce priveşte predicţia (sau preciziile lor aferente). În schimb, aceste semnale acţionează în noi pentru a coopta fluxurile de predicţii adecvate ce dezvăluie o lume de obiecte şi cauze distale”.
57. Dacă eşti alarmat de faptul că folosesc cuvântul „avatar”, aminteşte-ţi că tot ce percepi este virtual, inclusiv imaginea propriului corp. Ia în considerare ce am scris despre iluzia „schimbului de corpuri”, în Solms (2013), p. 15: „Subiectul conştiinţei se identifică cu corpul său extern (prezentarea obiectului) cam tot aşa cum un copil se proiectează pe sine în personajul animat pe care îl controlează într-un joc video. Reprezentarea este învestită rapid cu un simţ al sinelui, deşi nu e chiar sinele. Iată un experiment frapant ce ilustrează grăitor relaţia ilogică care există de fapt între sinele subiectiv şi corpul său extern. Petkova şi Ehrsson (2008) prezintă o serie de experimente de tip «schimb

de corpuri», în cadrul cărora în dreptul ochilor altor oameni sau ai unor manechine s-au amplasat camere video ce au transmis imagini din perspectivele respective către ochelari cu monitorizare video purtaţi de subiecţii experimentului, ceea ce le-a creat rapid subiecţilor iluzia că corpul celeilalte persoane sau al manechinului e propriul lor corp. Iluzia a fost atât de convingătoare, încât a persistat chiar şi atunci când subiecţii proiectaţi au dat mâna cu propriul corp. Existenţa iluziei a fost demonstrată obiectiv şi de faptul că, atunci când celălalt corp (propriu în mod iluzoriu) şi propriul corp (real) au fost ameninţate cu un cuţit, răspunsul de frică – „reacţia instinctivă” a corpului intern (măsurată în funcţie de frecvenţa cardiacă şi de conductanţa cutanată) – a fost mai puternic în cazul corpului iluzoriu [...]. Ne amintim astfel că cortexul nu e decât memorie cu acces aleatoriu”. Pentru această din urmă idee, vezi Ellis şi Solms (2018).

Capitolul 11. Problema dificilă

1. Davies (2019), pp. 184, 207.
2. Crick (1994). Citatul complet este dat la p. 249.
3. Chalmers (1995a), p. 201, sublinierea mea.
4. Crick (1994), p. 3, sublinierea mea.
5. Vezi Chalmers (1996), p. 251, sublinierea mea: „Cine ar fi crezut că bucata aceasta de materie cenuşie este ceea ce ar putea *să producă* experienţe subiective intense? Şi totuşi, le produce”.
6. Searle (1997), p. 28. Şi Searle se exprimă la fel: „Cum, mai exact, *cauzează* conştiinţa procesele neurobiologice din creier?” (1993, p. 3, sublinierea

mea).

7. Mă refer aici la închiderea cauzală a ceea ce e fizic: dacă conştiinţa nu este fizică sau dacă proprietăţile conştiente nu sunt proprietăţi fizice, atunci este greu de înţeles cum pot ele să influenţeze matricea cauzală a proceselor cerebrale. Vezi corespondenţa dintre René Descartes şi prinţesa Elisabeta a Boemiei (Shapiro, 2007).
8. Levine (1983). Ireductibilitatea experienţei fenomenologice la nişte procese fizice este numită şi „lacuna epistemică”.
9. Jackson (1982). Spun „cam aşa” pentru că am modificat puţin „argumentul cunoaşterii” adus de Jackson, nu numai ca să-l simplific, ci şi pentru că este inutil de crud în forma originală. În lumea reală, o asemenea cruzime ar afecta procesele psihologice descrise de Jackson.
10. Chalmers (2003), p. 104.
11. Vezi *Eseu asupra intelectului omenesc* de Locke (1690): „Este imposibil să concepem că materia, cu sau fără mişcare, a putut de la origine să aibă de la sine simţuri, percepere şi cunoaştere, de unde ar urma deci în chip evident că simţurile, perceperea şi cunoaşterea sunt însuşiri veşnic inseparabile de materie şi de fiecare particulă a ei” [John Locke, *Eseu asupra intelectului omenesc,* vol. II, traducere de Teodor Voiculescu, note de Dan Bădărău. Bucureşti: Editura Ştiinţifică, 1961, cartea IV, cap. X, §10, p. 234].
12. Nu vreau să sugerez că Locke a fost un

epifenomenalist. Există, desigur, alte poziţii dualiste, dar Jackson (1982), care a inventat exerciţiul de gândire despre Mary, a adoptat o poziţie epifenomenalistă. Mai târziu (1995) s-a răzgândit.

13. Oakley şi Halligan (2017). Aceşti autori atribuie, cel puţin, *o anumită* funcţie conştiinţei: capacitatea de a *informa despre* stările psihice (care, în sine, sunt inconştiente).
14. Ca să-l parafrazez pe Chalmers, e imaginabil din punct de vedere logic că „zombii filosofici” ar putea să emuleze toate funcţiile mecanice ale creierului fără să aibă vreo experienţă conştientă.
15. Libet *et al.* (1983). Totuşi, Libet nu crede că conştiinţa este epifenomenologică. El consideră că, în timpul celor 300 de milisecunde de dinaintea unei acţiuni, conştiinţa ar putea să aleagă să abandoneze acţiunea respectivă *(free won’t).* Acest tip de undă cerebrală se numeşte „potenţial premotor”. Studiile ulterioare sugerează că perioada de latenţă dintre potenţialul premotor şi decizia conştientă poate fi semnificativ mai mare de 300 de milisecunde.
16. Iar acest lucru poate fi explicat prin legea lui Hebb, care, din acelaşi motiv, explică şi faptul că experienţa ta trăită a amintirilor vechi de zece minute nu este atât de bine consolidată. Trebuie să precizez că toate acestea sunt în concordanţă cu ceea ce Chalmers (1995a) numeşte principiul
coerenţei structurale; vezi pp. 262-263.
17. Atkinson şi Shiffrin (1971).
18. Urmele mnezice pe termen scurt se degradează rapid din cauza mecanismelor de recaptare a

neurotransmiţătorilor ce readuc neuronii presinaptici în starea existentă înainte de formarea fiecărei urme, ceea ce le permite imediat să formeze alte urme. Vezi Mongillo, Barak şi Tsodyks (2008).

19. Chalmers (1995a), pp. 202-203. Chalmers explică sensul în care foloseşte termenul „funcţie”: „Aici, «funcţie» nu este folosit în sensul teleologic limitat, de ceva ce un sistem este conceput să facă, ci în sensul mai larg, de orice rol cauzal în generarea comportamentului pe care l-ar putea manifesta un sistem”.
20. *Ibidem,* pp. 204-205.
21. *Ibidem,* p. 205.
22. *Ibidem,* p. 204.
23. *Cf.* Nagel (1974): „Ca să apărăm fizicalismul, trăsăturile fenomenologice [ale conştiinţei] trebuie să aibă o descriere fizică. Dar, atunci când le analizăm caracterul subiectiv, pare că un asemenea rezultat e imposibil. Motivul e că fiecare fenomen subiectiv este conectat, practic, cu un singur punct de vedere şi pare inevitabil că o teorie fizică, obiectivă va abandona acest punct de vedere”. Searle (1997), p. 212, spune aşa: „Conştiinţa are o ontologie subiectivă sau la persoana întâi, aşa că nu poate fi redusă la ceva ce are ontologie subiectivă sau la persoana a treia. Dacă încerci să o reduci sau să o elimini pe una în favoarea celeilalte, vei lăsa ceva pe dinafară [...]. Nu poţi să reduci nici descărcarea neuronilor la sentimente, nici sentimentele la descărcarea neuronilor, deoarece în fiecare caz ar rămâne pe dinafară

obiectivitatea sau subiectivitatea aflată în ecuaţie”.

24. Chalmers (1995a), p. 203.
25. Havlik, Kozăkovă şi Horăce (2017).
26. Scrisoare din 23 decembrie 2017. Ne-am trimis articolul la *Journal of Consciousness Studies* pentru că acolo a publicat Chalmers (1995a) pentru prima dată problema dificilă.
27. Zahavi (2017) face această afirmaţie – este o ironie că o face împotriva afirmaţiilor lui Friston.
28. Cunosc foarte bine obiecţiile lui Searle în această privinţă, în sensul că qualia sunt o situaţie în care o singură realitate nu poate avea mai multe apariţii, deoarece ele *sunt* apariţiile (vezi nota 39, mai jos). Răspunsul meu pentru el ar fi: „Da, dar nu uita că apariţia *vizualizată* a activităţii corticale somato-senzoriale şi apariţia *simţită* a durerii fac parte din aceeaşi realitate resimţită”.
29. În această analogie, mă refer la două fenomene *exteroceptive:* percepţia fulgerului şi a tunetului. Când întreb ce le cauzează, aş putea să abordez întrebarea în două feluri. Aş putea fie să descriu evenimentele geofizice (electrice) care le generează, fie mecanismele senzoriale care consemnează (aceste două aspecte diferite legate de) evenimentele fizice. Am ales prima variantă în această analogie, lăsând a doua descriere pentru analiza problemei pe care o tratez în cartea de faţă, şi anume relaţia dintre evenimentele obiective şi cele subiective. Asta pentru că, după cum am văzut, *conştiinţa* ataşată evenimentelor senzoriale, indiferent dacă sunt activate exteroceptiv sau interoceptiv, este

întotdeauna endogenă. Această observaţie nu poate fi demonstrată în cadrul analogiei. Conştiinţa în sine, spun eu, nu este un semnal senzorial (exteroceptiv sau
interoceptiv), ci *sentimentul* semnalului.

30. În acest sens, te poţi folosi de optogenetică, de exemplu. Utilizând echipamente diferite, poţi *să asculţi* aşa-numitele *spike trains* ale semnalelor retiniene. După cum spune Wheeler, totul ţine doar de „răspunsurile evocate de echipament”.
31. Chalmers (1995a), pentru acest citat şi pentru următoarele, sublinierea mea *passim.*
32. La prima vedere, poziţia lui Chalmers pare să fie dualism al proprietăţilor clasic, în care mintea este o proprietate a materiei. Apoi, el concepe atât mintea, cât şi materia ca proprietăţi a altceva, numit „informaţie” (vezi mai jos). Asta ar putea să sune a monism al aspectului dual, dar al lui este tipul ciudat de informaţie pe care l-am discutat în capitolele anterioare: mintea este inerentă informaţiei, şi nu receptorului ei. La fel şi în cazul aspectului ce ţine de materie al informaţiei, după cum consideră Chalmers. Asta înseamnă că cele două aspecte ale informaţiei nu sunt proprietăţi în sensul epistemologic (nu sunt *apariţiile* a ceva numit informaţie), ci în sensul ontologic (sunt aspecte ale informaţiei *înseşi).* Totuşi, nu asta trebuie să ne preocupe. Ce mă preocupă cel mai mult este afirmaţia lui Chalmers că *orice* informaţie are un aspect psihic (ba chiar conştient). Aşadar, problema principală nu e dacă el consideră că aspectul psihic

al informaţiei este o proprietate sau o substanţă. Problema principală devine: este plauzibil să atribuim conştiinţă oricărei informaţii?

33. Nu uita că, pentru Chalmers, „caracteristicile fenomenologice ale lumii” exclud „ceea ce este fizic”.
34. Ar trebui să spun: echipamentul *care este* observatorul participant.
35. Am preluat această analogie din Hurley, Dennett şi Adams (2011).
36. Asta are legătură cu ce am spus mai sus (în capitolul 9) despre ipseitatea intenţională.
37. Chalmers (1995a), p. 217. Vezi şi afirmaţia lui anterioară: „Asta duce la o ipoteză firească – informaţia *(sau cel puţin unele informaţii)* are două aspecte de bază, unul fizic şi unul fenomenologic” (sublinierea mea).
38. Vezi capitolul 10, nota 14.
39. *Cf.* Searle (1992), p. 122: „Nu putem [distinge aparenţa de realitate] în cazul conştiinţei, deoarece conştiinţa constă din aparenţe. Când e vorba de aparenţă, nu putem face distincţia între aparenţă şi realitate pentru că aparenţa este realitatea”.
40. De asemenea, dacă Mary ar fi specialistă în corelativele neurale ale afectului, şi nu în cele ale vederii, nu ar fi posibil să nu aibă experienţe afective tot aşa cum nu are experienţe vizuale. Asta pentru că, dacă nu ar avea senzaţii, ar fi în comă (dacă nu moartă).
41. Vezi capitolul 5, nota 4.
42. Ne putem închipui deplasarea cu ajutorul unor roţi,

dar ceea ce s-a dezvoltat de fapt (în cazul nostru) au fost picioarele. Trebuie să fim atenţi să nu ridicăm ştacheta pentru explicaţiile referitoare la conştiinţă mai mult decât cele referitoare la orice altceva din biologie.

43. Chalmers (1995a), pp. 203-204.

Capitolul 12. Cum se face o minte

1. Dar o caracatiţă s-ar putea să nu fie de acord. (Sistemul lobar vertical pare să fie cel mai apropiat corespondent la cefalopode al paliumului vertebratelor.)
2. La început, din grup am făcut parte Tristan Hromnik, Jonathan Shock şi eu, apoi grupul a inclus şi alţi fizicieni, informaticieni şi specialişti în bioinginerie medicală – George Ellis, Rowan Hodson, Leen Remmelzwaal, Amit Mishra, Dean Rance, Dawie van den Heever şi Julianne Blignaut –, precum şi pe neuropsihologii Joshua Martin, Aimee Dollman şi Donne van der Westhuizen. Echipa continuă să crească, deşi Hromnik nu mai face parte din ea, iar Martin s-a mutat la Berlin.
3. Searle (1980). Damasio (2018) a avut o opinie similară.
4. Mă refer la argumentul „qualia dansante” al lui Chalmers, ce

   are ca punct de plecare argumentul „spectrelor inversate” al lui Locke. Vezi Chalmers (1995a, 1995b, 2011).
5. Chalmers (1995a), pp. 214-215.

6. *Ibidem,* p. 215, sublinierea mea.
7. *Ibidem,* sublinierea mea.
8. Am avut lungi discuţii cu el în această privinţă şi impresia mea e că măcar are o atitudine deschisă.
9. Ethier *et al.* (2012), Hochberg *et al.* (2012), Collinger *et al.* (2013), Bouton *et al.* (2016), Capogrosso *et al.* (2016).
10. Capogrosso *et al.* (2016), p. 284.
11. *Ibidem.*
12. Abu-Hassan *et al.* (2019).
13. Pasley *et al.* (2012).
14. Nishimoto *et al.* (2011).
15. Horikawa *et al.* (2013).
16. Herff *et al.* (2015).
17. Inclusiv cele de sus în jos, de tipul discutat în cadrul temei „gândire” din capitolul 10 (precum imaginarea şi visul).
18. *Cf.* Kurzweil (2005).
19. Solms şi Turnbull (2002), pp. 70-71, a doua subliniere îmi aparţine.
20. Vezi Solms (1996, 1997b).
21. Renumitul fizician Richard Feynman a avut aceeaşi perspectivă asupra înţelegerii mecaniciste în general. Nu vreau să spun că conştiinţa obţinută prin inginerie inversă rezolvă problema. E posibil să montezi ceva microcip cu microcip fără să te gândeşti la asta şi fără să înţelegi cevaul. Vreau să spun însă că, *dacă* îl înţelegi, atunci ar trebui să-l poţi crea prin inginerie inversă.
22. Vezi Reggia (2013), pentru o trecere în revistă a cercetărilor anterioare în acest sens.

23. Întrucât abordarea învăţării prin consolidare presupune un criteriu de obiectiv, acesta va fi: *supravieţuirea sistemului în medii neprevăzute.*
24. Ca să fie clar: întruparea poate fi simulată (ca şi toţi parametrii fizici descrişi mai jos). Din punctul de vedere al sistemului, nu contează ce se întâmplă de fapt „în exterior”, ci numai ce se întâmplă în model în legătură cu *informaţiile pe care le primeşte* din exterior. Prin urmare, echipa de cercetare poate să simuleze un mediu din care sistemul să-şi facă un model şi vom proceda astfel pentru primele generaţii ale sistemului pe care îl propunem. Dacă am proceda altfel, ar fi nevoie de extrem de mult timp şi, sincer, ar fi periculos (de exemplu, gândeşte-te la parametrul de supraîncălzire, descris mai jos). Când spun că este mai „realist” să întrupăm sistemul din punct de vedere fizic, mă refer la faptul că, odată cu mişcarea fizică (de exemplu), apar probleme de modelare ce nu apar în cazul mişcării simulate, iar acest lucru s-ar putea dovedi a fi important pentru un sistem cu adevărat similar unuia însufleţit. Din acest motiv şi din altele, generaţiile ulterioare ale sistemului pe care îl propunem vor fi întrupate în roboţi.
25. Adică vom codifica reflexe şi instincte artificiale. Vom folosi şi „algoritmi genetici”.
26. De pe Wikipedia (https://en.wikipedia.org/wiki/Artificial_consciousness), 21 martie 2020. Pentru o opinie alternativă, vezi Reggia (2013): „Autorul acestei recenzii crede că niciunul dintre studiile anterioare

examinate, chiar dacă se afirmă altceva, nu oferă încă vreun argument convingător cu privire la modul cum va duce abordarea studiată în cele din urmă la o conştiinţă artificială concretizată. Totodată şi pe un ton mai marcat pozitiv, încă nu s-au adus dovezi (inclusiv în studiile analizate în această recenzie) că, în cele din urmă, conştiinţa automată concretizată nu ar fi posibilă, opinie exprimată şi de alţii".

27. Solms şi Turnbull (2002), pp. 68-69.
28. Vezi Colby, Watt şi Gilbert (1966), Weizenbaum (1976). Un program de calculator numit „Eugene Goostman", care simulează un băiat ucrainean de 13 ani, a trecut testul Turing în cadrul unui eveniment organizat în 2014 la Royal Society of London.
29. De exemplu, după cum se ştie, calculatoarele îi pot bate şi pe cei mai buni jucători umani şi de şah, şi de go, care este mai dificil decât şahul.
30. Vezi Haikonen (2012).
31. Este emoţionant, având în vedere cât a suferit chiar Alan Turing (creatorul testului) din cauza prejudecăţilor.
32. Acum mai mulţi ani, am spus în cadrul unei reuniuni desfăşurate la Viena între psihanalişti şi ingineri specialişti în inteligenţă artificială că existenţa conştiinţei artificiale ar putea fi demonstrată, de exemplu, dacă s-ar găsi dovezi de psihopatologie artificială: „În măsura în care inginerii reuşesc să imite cu acurateţe mintea umană, vor descoperi că modelul lor este predispus la anumite tipuri de disfuncţionalitate. Aproape că suntem tentaţi să

folosim acest lucru drept criteriu al succesului lor” (Solms, 2008).

33. Mathur, Lau şi Guo (2011).
34. Gândeşte-te la cazul extrem al „corporaţiilor autonome descentralizate”.
35. Vezi Lin, Abney şi Bekey (2011).
36. Bill Gates, Stephen Hawking şi Elon Musk, de exemplu, şi-au exprimat rezerve serioase pe această temă.
37. Nu pretind că această convingere, singură, constituie o justificare etică pentru un asemenea demers. Faptul că cineva va comite o crimă undeva, la un moment dat, nu justifică o crimă a mea aici şi acum. Te rog să citeşti mai departe...
38. Solms şi Friston (2018), Solms (2019a), Solms (2020b).
39. Iată o listă parţială: „Where does consciousness fit in the Bayesian brain?”, al 18-lea Congres Internaţional de Neuropsihanaliză, University College London, 2017; „How and why consciousness arises”, departamentul de fizică, University of Cape Town, 2017; „How and why consciousness arises”, Centrul de studii asupra subiectivităţii, Kobenhavns Universitet, Copenhaga, 2017; „The self aş feeling and memory”, Ruhr Universität, Bochum, 2018; „The conscious id, the psychoanalytic process and the hard problem of consciousness”, departamentul de filosofie, New York University, 2019; „Why and how consciousness arises”, departamentul de psihiatrie, spitalul Mount Sinai, New York, 2019; „Why are we conscious? Lessons from neuroscience”, University of

Vermont College of Medicine, Burlington, 2019; „What is consciousness?”, The Melbourne Brain Centre, Australia, 2019; „Consciousness itself”, The Science of Consciousness, Interlaken, Elveţia, 2019; „Consciousness itself is affect”, Hochschule für Philosophie, München, Burkardus Haus, Würzburg, Germania, 2019; „The hard problem of consciousness”, departamentul de filosofie, University of Cape Town, 2019; „Consciousness is predictive work in progress”, spitalul Ichilov, Tel Aviv, 2019; „Why and how consciousness arises”, Italian Psychoanalytic Dialogues, Roma, 2020.

40. Havlik, Kozakova şi Horace (2017).
41. Carhart-Harris şi Friston (2010), Carhart-Harris *et al.* (2014), Carhart-Harris (2018).
42. Vezi capitolul 9, nota 16.
43. www.linking-ai-principles.org/term/656. Vezi The Public Voice Coalition (2018).

Post-scriptum

1. Legea afectului: „Dacă un comportament este însoţit constant
de plăcere, se va intensifica, iar dacă este însoţit constat de neplăcere, va scădea”.
2. Legea lui Friston: „Toate cantităţile care se pot schimba, cu alte cuvinte, care fac parte din sistem, se vor schimba cu siguranţă pentru a minimiza energia liberă”.

Bibliografie

Abbott, A. (2020), What animals really think. *Nature,* 584: 182185.

Absher, J. şi Benson, D. (1993), Disconnection syndromes: an overview of Geschwind's contributions. *Neurology,* 43: 862867.

Abu-Hassan, K., Taylor, J., Morris, P. *et al.* (2019), Optimal solid state neurons. *Nature Communications,* 10: 5309.

Adams, R., Shipp, S. şi Friston, K. (2013), Predictions not commands: active inference in the motor system. *Brain Structure and Function,* 218: 611-643.

Addis, D., Wong, A. şi Schacter, D. (2007), Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia,* 45: 1363-1377.

Ainley, V., Apps, M.A.J., Fotopoulou, A. şi Tsakiris, M. (2016), „Bodily precision": a predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society of London, B,* 371: (1708), doi.org/10.1098/rstb.2016.0003

Alboni, P. (2014), Vasovagal syncope aş a manifestation of an evolutionary selected trait. *Journal of Atrial Fibrillation,* 7: 1035.

Aserinsky, E. şi Kleitman, N. (1953), Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science,* 118: 273-274.

Ashby, W. (1947), Principles of the self-organizing dynamic system. *Journal of General Psychology,* 37: 125-128.

Atkinson, R. şi Shiffrin, R. (1971), The control of short-term memory. *Scientific American,* 225: 82-90.

Baars, B. (1989), *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press.

Baars, B. (1997), *În the Theatre of Consciousness.* Oxford: Oxford University Press.

Bailey, P. şi Davis, E. (1942), The syndrome of obstinate progression in the cat. *Experimental Biology and Medicine,* 52: 307.

Bargh, J. şi Chartrand, T. (1999), The unbearable automaticity of being. *American Psychologist,* 54: 462-479.

Barrett, L.F. (2017), *How Emotions are Made: The Secret Life of the Brain.* New York: Houghton Mifflin Harcourt.

Bastos, A., Usrey, W., Adams, R. *et al.* (2012), Canonical microcircuits for predictive coding. *Neuron,* 76: 695-711.

Bastos, A., Vezoli, J., Bosman, C. *et al.* (2015), Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron,* 85:

390-401.
Bayes, T. (1763), An essay towards solving a problem in the doctrine of chances. [Communicated by Mr. Price, in a letter to John Canton.] *Philosophical Transactions of the Royal Society of London,* 53: 370-418.
Bechtel, W. şi Richardson, R. (1998), Vitalism. În E. Craig (ed.), *Routledge Encyclopedia of Philosophy*, 9. Londra: Routledge, pp. 639-643.
Bentley, B., Branicky, R., Barnes, C. *et al.* (2016), The multilayer connectome of *Caenorhabditis elegans. PLoS Computational Biology,* 12: e1005283, doi.org/10.1371/journal.pcbi.1005283
Berlin, H. (2011), The neural basis of the dynamic unconscious. *Neuropsychoanalysis,* 13: 5-31.
Berlin, H. (2013), The brainstem begs the question: „petitio principii". *Neuropsychoanalysis,* 15: 25-29.
Berridge, K. (2003), Pleasures of the brain. *Brain and Cognition,* 52: 106-128.
Besharati, S., Forkel, S.J., Kopelman, M., Solms, M., Jenkinson, P.M. şi Fotopoulou, A. (2014), The affective modulation of motor awareness in anosognosia for hemiplegia: behavioural and lesion evidence. *Cortex,* 61: 127-140.
Besharati, S., Forkel, S., Kopelman, M., Solms, M., Jenkinson, P. şi Fotopoulou, A. (2016), Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia. *Brain*, 139: 971-985.
Besharati, S., Fotopoulou, A. şi Kopelman, M. (2014), What is it like to be confabulating? În A.L. Mishara,

A. Kranjec, P. Corlett, P. Fletcher şi M.A. Schwartz (eds.), *Phenomenological Neuropsychiatry: How Patient Experience Bridges Clinic with Clinical Neuroscience.* New York: Springer.

Bienenstock, E., Cooper L. şi Munro P. (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2: 32-48.

Blake, Y., Terburg, D., Balchin, R., van Honk, J. şi Solms, M. (2019), The role of the basolateral amygdala in dreaming, *Cortex*, 113: 169-183, doi.org/10.1016/j.cortex.2018.12.016

Block, N. (1995), On a confusion about a function of consciousness. *Behavioral and Brain Sciences,* 18: 227-287.

Blomstedt, P., Hariz, M., Lees, A. *et al.* (2008), Acute severe depression induced by intraoperative stimulation of the substantia nigra: a case report. *Parkinsonism and Related Disorders,* 14: 253-156.

Bogen, J. (1995), On the neurophysiology of consciousness: 1. An overview. *Consciousness and Cognition,* 4: 52-62.

Bouton, C., Shaikhouni, A., Annetta, N. *et al.* (2016), Restoring cortical control of functional movement in a human with quadriplegia. *Nature,* 533: 247-250.

Bowlby, J. (1969), *Attachment.* Londra: Hogarth Press.

Braun, A. (1999), The new neuropsychology of sleep. *Neuropsychoanalysis,* 2: 196-201.

Braun, A., Balkin, T., Wesenten, N. *et al.* (1997), Regional cerebral blood flow throughout the sleep-

wake cycle. An H2(15)O PET study. *Brain,* 120: 1173-1197.

Brentano, F. (1874), *Psychologie vom empirischen Standpunkte.* Leipzig: Duncker and Humbolt.

Broca, P. (1861), Sur le principe des localisations cerebrales. *Bulletin de la Societe d'Anthropologie,* 2: 190-204.

Broca, P. (1865), Sur le siege de la faculte du langage articule. *Bulletin de la Societe d'Anthropologie,* 6: 377-373.

Brown, H., Adams, R., Parees, I., Edwards, M. şi Friston, K. (2013), Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14: 411-427.

Cameron-Dow, C. (2012), Do dreams protect sleep? Testing the Freudian hypothesis of the function of dreams. MA dissertation, University of Cape Town.

Campbell, A. (1904), Histological studies on the localisation of cerebral function. *Journal of Mental Science,* 50: 651-659.

Capogrosso, M., Milekovic, T., Borton, D. *et al.* (2016), A brainspine interface alleviating gait deficits after spinal cord injury in primates. *Nature,* 539: 284-288.

Carhart-Harris, R. (2018), The entropic brain – revisited. *Neuropharmacology,* 142: 167-178.

Carhart-Harris, R. şi Friston, K. (2010), The default-mode, egofunctions and free-energy: a neurobiological account of Freudian ideas. *Brain,* 133: 1265-1283.

Carhart-Harris, R., Leech, R., Hellyer, P. *et al.* (2014), The entropic brain: a theory of conscious states

informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience,* 8: Article 20.

Chabris, C. şi Simons, D. (2010), *The Invisible Gorilla: and Other Ways Our Intuitions Deceive Us.* Londra: Crown Publishers/Random House.

Chalmers, D. (1995a), Facing up to the problem of consciousness. *Journal of Consciousness Studies,* 2: 200-219.

Chalmers, D. (1995b), Absent qualia, fading qualia, dancing qualia. În T. Metzinger (ed.), *Conscious Experience.* Paderborn: Ferdinand Schoningh, pp. 309-328.

Chalmers, D. (1996), *The Conscious Mind: În Search of a Fundamental Theory.* New York: Oxford University Press.

Chalmers, D. (2003), Consciousness and its place in nature. În S. Stich şi T. Warfield (eds.), *Blackwell Guide to the Philosophy of Mind*. Londra: Blackwell, pp. 102-142.

Chalmers, D. (2011), A computational foundation for the study of cognition. *Journal of Cognitive Science,* 12: 325-359.

Charcot J.-M. (1883), Un cas de suppression brusque et isolee de la vision mentale des signes et des objets (formes et couleurs). *Progres Medical*, 11: 568.

Chew, Y., Tanizawa, Y., Cho, Y. *et al.* (2018), An afferent neuropeptide system transmits mechanosensory signals triggering sensitization and arousal in *C. elegans. Neuron,* 99: 1233-1246.

Cisek, P. şi Kalaska, J. (2010), Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience,* 33: 269-298.
Claparede, E. (1911), Recognition et moitié. *Archives of Psychology Geneve,* 11: 79-90.
Clark, A. (2015), *Surfing Uncertainty: Prediction, Action, and the Embodied Mind.* New York: Oxford University Press.
Clark, A. (2017), Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Nous*, 51: 727753.
Coenen, A. (2007), Consciousness without a cortex, but what kind of consciousness is this? *Behavioral and Brain Sciences,* 30: 87-88.
Coenen, V., Bewernick, B., Kayser, S. *et al.* (2019), Superolateral medial forebrain bundle deep brain stimulation in major depression: a gateway trial. *Neuropsychopharmacology,* 44: 1224-1232, doi.org/10.1038/s41386-019-0369-9
Colby, K., Watt, J. şi Gilbert, J. (1966), A computer method of psychotherapy. *Journal of Nervous and Mental Disease,* 142: 148-152.
Cole, S., Fotopoulou, A., Oddy, M. şi Moulin, C. (2014), Implausible future events in a confabulating patient with an anterior communicating artery aneurysm. *Neurocase*, 20: 208-224.
Collinger J., Wodlinger, B., Downey, J. *et al.* (2013), Highperformance neuroprosthetic control by an individual with
tetraplegia. *The Lancet,* 381: 557-564.

Coltheart, M. şi Turner, M. (2009), Confabulation and delusion. În W. Hirstein (ed.), *Confabulation: Views from Neuroscience, Psychiatry, Psychology and Philosophy.* New York: Oxford University Press, p. 173.

Conant, R. şi Ashby, W. (1970), Every good regulator of a system must be a model of that system. *International Journal of Systems Science,* 1: 89-97.

Coren, S. şi Porac, C. (1974), The fading of stabilized images: eye movements and information processing. *Perception & Psychophysics,* 16: 529-534.

Corlett, P. şi Fletcher, P. (2014), Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry*, 1: 399-402.

Craig, A.D. (2009), How do you feel – now? The anterior insula and human awareness. *Nature Reviews Neuroscience,* 10: 5970.

Craig, A.D. (2011), Significance of the insula for the evolution of human awareness of feelings from the body. *Annals of the New York Academy of Sciences,* 1225: 72-82.

Crick, F. (1994), *The Astonishing Hypothesis: The Scientific Search for the Soul.* New York: Charles Scribner's Sons.

Crick, F. (2004), Foreword. În C. Koch, *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company.

Crick, F. şi Koch, C. (1990), Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2: 263-275.

Crucianelli, L., Krahe, C., Jenkinson, P. şi Fotopoulou, A. (2017), Interoceptive ingredients of body ownership: affective touch and cardiac awareness in the rubber hand illusion. *Cortex,* 104: 180-192, doi.org/10.1016/j.cortex.2017.04.018

Cukur, T., Nishimoto, S., Huth, A. şi Gallant J. (2013), Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience,* 16: 763-770.

Dahan, L., Astier, B., Vautrelle, N. *et al.* (2007), Prominent burst firing of dopaminergic neurons in the ventral tegmental area during paradoxical sleep. *Neuropsychopharmacology,* 32: 1232-1241.

Damasio, A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: Putnam.

Damasio, A. (2018), *The Strange Order of Things: Life, Feeling, and the Making of Cultures.* Londra: Penguin Random House.

Damasio, A. şi Carvalho, G. (2013), The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience,* 14: 143-152.

Damasio, A. şi Damasio, H. (1989), *Lesion Analysis in Neuropsychology.* New York: Oxford University Press.

Damasio, A., Damasio, H. şi Tranel, D. (2013), Persistence of feelings and sentience after bilateral damage of the insula. *Cerebral Cortex,* 23: 833-846.

Damasio, A., Grabowski, T., Bechara, A. *et al.* (2000), Subcortical and cortical brain activity during the feeling of selfgenerated emotions. *Nature Neuroscience*, 3: 1049-1056.

Darwin, C. (1859), *On the Origin of Species*. Londra: John Murray.

Darwin, C. (1872), *The Expression of Emotions in Man and Animals*. Londra: John Murray.

Davies, P. (2019), *The Demon in the Machine: How Hidden Webs of Information are Solving the Mystery of Life.* Londra: Allen Lane.

Debiec, J., Doyere, V., Nader, K. şi LeDoux, J. (2006), Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala. *Proceedings of the National Academy of Sciences,* 103: 3428-3433.

Decety, J. şi Fotopoulou, A. (2015), Why empathy has a beneficial impact on others in medicine: unifying theories. *Frontiers in Behavioral Neuroscience,* 8: 457.

Dehaene, S. şi Changeux, J.-P. (2005), Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biology,* 3: e141.

Dehaene, S. şi Changeux, J.-P. (2011), Experimental and theoretical approaches to conscious processing. *Neuron*, 70: 200-227.

Dehaene, S. şi Naccache, L. (2001), Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition,* 79: l-37.

Dement, W. şi Kleitman, N. (1957), The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *Journal of Experimental Psychology,* 53: 339-346.

Depaulis, A. şi Bandler, R. (1991), *The Midbrain Periaqueductal Gray Matter: Functional, Anatomical, and Neurochemical Organization.* New York: Plenum Press.

Ditchburn, R. şi Ginsborg, B. (1952), Vision with a stabilized retinal image, *Nature*, 170: 36-37.

Domhoff, W. (2017), *The Emergence of Dreaming: Mind-Wandering, Embodied Simulation, and the Default Network.* New York: Oxford University Press.

Du Bois-Reymond, E. (1848-1884), *Untersuchungen uber thierische Electricitdt,* 2. Berlin: Reimer.

Du Bois-Reymond, E., ed. (1918), *Jugendbriefe von Emil Du Bois- Reymond an Eduard Hallmann, zu seinem hundertsten Geburtstag, dem 7. November 1918.* Berlin: Reimer.

Dudai, Y. (2000), The shaky trace. *Nature,* 406: 686-687.

Edelman, G. (1990), *The Remembered Present: A Biological Theory of Consciousness.* New York: Basic Books.

Edlow, B., Takahashi, E., Wu, O. *et al.* (2012), Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. *Journal of Neuropathology and Experimental Neurology,* 71: 531-546.

Einstein, A. (1905), Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik,* 17: 132-148.

Eisenberger, N. (2012), The neural bases of social pain: evidence for shared representations with physical pain. *Psychosomatic Medicine,* 74: 126-135.

Ekman, P., Friesen, W., O'Sullivan, M. *et al.* (1987), Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53: 712-717.

Ellis, G. şi Solms, M. (2018), *Beyond Evolutionary Psychology: How and Why Neuropsychological Modules Arise.* Cambridge: Cambridge University Press.

England, J. (2013), Statistical physics of self-replication. *Journal of Chemical Physics,* 139: 121923, doi.org/10.1063/1.4818538

Ethier, C., Oby, E., Bauman, M. şi Miller, L. (2012), Restoration of grasp following paralysis through brain-controlled stimulation of muscles. *Nature,* 485: 368-371.

Ezra, M., Faull, O., Jbabdi, S. şi Pattinson, K. (2015), Connectivitybased segmentation of the periaqueductal gray matter in human with brainstem optimized diffusion MRI. *Human Brain Mapping,* 36: 3459-3471.

Feldman, H. şi Friston, K.J. (2010), Attention, uncertainty, and freeenergy. *Frontiers in Human Neuroscience,* 4: 215, doi.org/10.3389/fnhum.2010.00215

Ferrarelli, F. şi Tononi, G. (2011), The thalamic reticular nucleus and schizophrenia. *Schizophrenia Bulletin,* 37: 306-315.

Fischer, D., Boes, A., Demertzi, A. *et al.* (2016), A human brain network derived from coma-causing brainstem lesions. *Neurology,* 87: 2427-2434.

Flechsig, P. (1901), Developmental (mylogenetic) localisation of the cerebral cortex in the human subject. *The Lancet,* 2: 1027-1029.

Flechsig, P. (1905), Gehirnphsyiologie und Willenstheorien. *Fifth International Psychology Congress,* Rome, pp. 73-89. În G. von Bonin (ed.), *Some Papers on the Cerebral Cortex.* Springfield, IL: Charles C. Thomas, pp. 181-200.

Forrester, G., Davis, R., Mareschal, D. *et al.* (2018), The left cradling bias: an evolutionary facilitator of social cognition? *Cortex*, 118: 116-131, doi.org/10.1016/j.cortex.2018.05.011

Fotopoulou, A. (2008), False-selves in neuropsychological rehabilitation: the challenge of confabulation. *Neuropsychological Rehabilitation,* 18: 541-565.

Fotopoulou, A. (2009), Disentangling the motivational theories of confabulation. În W. Histein (ed.), *Confabulation: Views from Neurology, Psychiatry, and Philosophy.* New York: Oxford University Press.

Fotopoulou, A. (2010a), The affective neuropsychology of confabulation and delusion. *Cognitive Neuropsychiatry,* 15: 38-63.

Fotopoulou, A. (2010b), The affective neuropsychology of confabulation and delusion. În R. Langdon şi M. Turner (eds.), *Confabulation and Delusion,* New York: Psychology Press, pp. 38-63.

Fotopoulou, A. (2013), Beyond the reward principle: consciousness aş precision seeking. *Neuropsychoanalysis,* 15: 33-38.

Fotopoulou, A. şi Conway, M. (2004), Confabulation pleasant and unpleasant. *Neuropsychoanalysis,* 6: 26-33.
Fotopoulou, A., Conway, M., Birchall, D., Griffiths, P. şi Tyrer, S. (2007), Confabulation: revising the motivational hypothesis. *Neurocase*, 13: 6-15.
Fotopoulou, A., Conway, M. şi Solms, M. (2007), Confabulation: motivated reality monitoring. *Neuropsychologia,* 45: 21802190.
Fotopoulou, A., Conway, M., Solms, M., Kopelman, M. şi Tyrer, S. (2008a), Self-serving confabulation in prose recall. *Neuropsychologia*, 46: 1429-1441.
Fotopoulou, A., Conway, M., Tyrer, S., Birchall, D., Griffiths, P. şi Solms, M. (2008b), Is the content of confabulation positive? An experimental study. *Cortex,* 44: 764-772.
Fotopoulou, A., Solms, M. şi Turnbull, O. (2004), Wishful reality distortions in confabulation: a case report. *Neuropsychologia*, 42: 727-744.
Fotopoulou, A. şi Tsakiris, M. (2017), Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychoanalysis*, 19: 3-76.
Frank, J. (1946), Clinical survey and results of 200 cases of prefrontal leucotomy. *Journal of Mental Sciences,* 92: 497508.
Frank, J. (1950), Some aspects of lobotomy (prefrontal leucotomy) under psychoanalytic scrutiny. *Psychiatry,* 13: 45-52.
Frank, M. (2005), Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of

cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience,* 1: 51-72.
Freud, S. (1883), Einleitung in der Nervenpathologie. Unpublished book manuscript. Washington, DC: Library of Congress.
Freud, S. (1886), Report on my studies in Paris and Berlin. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 1. Londra: Hogarth, pp. l-15.
Freud, S. (1888), Gehirn. I. Anatomie des Gehirns. În A. Villaret (ed.), *Handwörterbuch der gesamten Medizin,* 1. Stuttgart: Ferdinand Enke, pp. 684-691.
Freud, S. (1891), *On Aphasia.* New York: International Universities Press.
Freud, S. (1893a), Charcot. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 3. Londra: Hogarth, pp. 11-23.
Freud, S. (1893b), Some points for a comparative study of organic and hysterical motor paralyses. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 155-172.
Freud, S. (1894), The neuro-psychoses of defence. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 3. Londra: Hogarth, pp. 45-61.
Freud, S. (1895), Studies on hysteria. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 2. Londra: Hogarth [ed.rom.: *Studii despre isterie,* traducere din limba germană de Maria şi Ion

Nastasia, cuvânt înainte de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2014].
Freud, S. (1900), The interpretation of dreams. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 4 and 5. Londra: Hogarth [ed.rom.: *Opere esenţiale,* vol. 2: *Interpretarea viselor,* traducere din limba germană de Roxana Melnicu. Bucureşti: Trei, 2010].
Freud, S. (1901), The psychopathology of everyday life. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 6. Londra: Hogarth [ed.rom.: *Psihopatologia vieţii cotidiene. Despre uitare, greşeli de vorbire, superstiţie şi eroare,* traducere de Herta Spuhn şi Daniela Ştefănescu, ediţia a Il-a. Bucureşti: Trei, 2010, e-book].
Freud, S. (1912), A note on the unconscious in psycho-analysis. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 12. Londra: Hogarth, pp. 255-266.
Freud, S. (1914), On narcissism: an introduction. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 14. Londra: Hogarth, pp. 67-102 [ed. rom.: „Pentru a introduce narcisismul”, traducere de Gilbert Lepădatu, în *Opere esenţiale,* vol. 3: *Psihologia inconştientului,* notă asupra ediţiei de Raluca Hurduc, note introductive de Roxana Melnicu. Bucureşti: Trei, 2017, pp. 35-68].
Freud, S. (1915a), Instincts and their vicissitudes. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 14. Londra: Hogarth, pp.

117-140 [ed.rom.: „Pulsiuni şi destine ale pulsiunilor”, traducere de Gilbert Lepădatu, în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, notă asupra ediţiei de Raluca Hurduc, note introductive de Roxana Melnicu. Bucureşti: Trei, 2017, pp. 71-96].

Freud, S. (1915b), The unconscious. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 14. Londra: Hogarth, pp. 166-204.

Freud, S. (1920), Beyond the pleasure principle. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 18. Londra: Hogarth, pp. 7-64 [ed.rom.; „Dincolo de principiul plăcerii”, în *Opere esenţiale,* vol. 3: *Psihologia inconştientului,* traducere de G. Purdea şi Vasile Dem. Zamfirescu. Bucureşti: Trei, 2017, pp. 213-286].

Freud, S. (1923), The ego and the id. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 19. Londra: Hogarth, pp. 12-59.

Freud, S. (1925), A note upon „the mystic writing-pad”. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 16. Londra: Hogarth, pp. 227-232.

Freud, S. (1940 [1939]), An outline of psycho-analysis. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 23. Londra: Hogarth, pp. 144-207 [ed.rom.: „Compendiu de psihanaliză”, în *Despre psihanaliză,* traducere de Daniela Ştefănescu, prefaţă de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2014, pp. 169-253].

Freud, S. (1950a [1895]), Extracts from the Fliess

papers. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 177-280.
Freud, S. (1950b [1895]), Project for a scientific psychology. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 283-397 [ed.rom.: *Opere,* vol. XVII: *Proiect de psihologie,* traducere din limba germană de Daniela Ştefănescu, cuvânt înainte de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2007].
Freud, S. (1994 [1929]), Letter to Einstein, 1929. În I. Grubrich- Simitis (1995), „No greater, richer, more mysterious subject ... than the life of the mind". *International Journal of Psychoanalysis,* 76: 115-122.
Friston, K. (2005), A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, B,* 360: 815-836.
Friston, K. (2009), The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences,* 13: 293-301.
Friston, K. (2013), Life aş we know it. *Journal of the Royal Society Interface,* 10: 20130475, doi.org/10.1098/rsif.2013.0475
Friston, K., Breakspear, M. şi Deco, G. (2012), Perception and selforganized instability. *Frontiers in Computational Neuroscience,* 6: 44.
Friston, K., Rigoli, F., Ognibene, D. *et al.* (2015), Active inference and epistemic value. *Cognitive Neuroscience,* 6: 187-214.
Friston, K., Schwartenbeck, P., FitzGerald, T.,

Moutoussis, M., Behrens, T. şi Dolan, R. (2014), The anatomy of choice: dopamine and decision-making. *Philosophical Transactions of the Royal Society of London, B*, 369: doi.org/10.1098/rstb.2013.0481

Friston, K., Shiner, T., Fitzgerald, T., Galea, J., Adams, R., Brown, H., Dolan, R., Moran, R., Stephan, K. şi Bestmann, S. (2012), Dopamine, affordance and active inference. *PLoS Computational Biology*, 8: e1002327

Friston, K. şi Stephan, K. (2007), Free-energy and the brain. *Synthese*, 159: 417-458.

Friston, K., Stephan, K., Montague, R. şi Dolan, R. (2014), Computational psychiatry: the brain aş a phantastic organ. *Lancet Psychiatry*, 1: 148-158.

Frith, C., Blakemore, S. şi Wolpert, D. (2000), Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London, B,* 355: 17711788.

Galin, D. (1974), Implications for psychiatry of left and right cerebral specialization: a neurophysiological context for unconscious processes. *Archives of General Psychiatry,* 31: 572-583.

Garcia-Rill, E. (2017), Bottom-up gamma and stages of waking. *Medical Hypotheses,* 104: 58-62.

Gloor, P. (1992), Role of the amygdala in temporal lobe epilepsy. În J. Aggleton (ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction.* New York: Wiley- Liss, pp. 505-538.

Golaszewski, S. (2016), Coma-causing brainstem lesions. *Neurology*, 87: 10.

Goodglass, H. (1986), Norman Geschwind (1926-1984). *Cortex,* 22: 7-10.

Gosseries, O., Schnakers, C., Ledoux, D. *et al.* (2011), Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Functional Neurology,* 26: 25-30.

Gregory, R. (1980), Perceptions aş hypotheses. *Philosophical Transactions of the Royal Society of London, B,* 290: 181-197.

Haikonen, P. (2012), *Consciousness and Robot Sentience*. New Jersey: World Scientific.

Harding, D. (1961), *On Having No Head.* Londra: Sholland Trust.

Harlow, J. (1868), Passage of an iron rod through the head. *Boston Medical and Surgical Journal,* 39: 389-393.

Hartmann. E., Russ. D., Oldfield. M., Falke. R. şi Skoff. B. (1980),

Dream content: effects of l-DOPA. *Sleep Research,* 9: 153.

Hassin, R., Bargh, J., Engell, A. şi McCulloch, K. (2009), Implicit working memory. *Consciousness and Cognition,* 18: 665-678.

Havlik, M., Kozakova, E. şi Horace, J. (2017), Why and how: the future of the central questions of consciousness. *Frontiers in Psychology,* 8: 1797, doi.org/10.3389/fpsyg.2017.01797

Hebb, D. (1949), *The Organization of Behavior: A Neuropsychological Theory.* New York: Wiley.
Helmholtz, H. von (1867), *Handbuch der physiologischen Optik,* 3. Leipzig: Voss.
Helmholtz, H. von (1892), Goethes Vorahnungen kommender naturwissenschaftlicher Ideen. În *Vorträge und Reden,* 2. Braunschweig: Friedrich Vieweg und Sohn, pp. 335-361.
Herff, C., Heger, D., De Pesters, A. *et al.* (2015), Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience*, 9: 217, doi.org/10.3389/fnins.2015.00217
Hering, E. (1879), Der Raumsinn und die Bewegungen des Auges. În L. Hermann (ed.), *Handbuch der Physiologie*, 3. Part 1: *Physiologie des Gesichtssinnes.* Leipzig: Vogel, pp. 343-601.
Hesselmann, G., Sadaghiani, S., Friston, K. şi Kleinschmidt, A. (2010), Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS One,* 5(3): e9926, doi.org/10.1371/journal.pone.0009926
Hobson, J.A. (2009), REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience,* 10: 803813.
Hobson, J.A. şi Friston (2012), Waking and dreaming consciousness: neurobiological and functional considerations. *Progress in Neurobiology*, 98: 82-98.
Hobson, J.A. şi Friston, K. (2014), Consciousness,

dreams, and inference: the Cartesian theatre revisited. *Journal of Consciousness Studies,* 21: 6-32.
Hobson, J.A. şi McCarley, R. (1977), The brain aş a dream state generator: an activation-synthesis hypothesis of the dream process. *American Journal of Psychiatry,* 134: 1335-1348.
Hobson, J.A., McCarley, R. şi Wyzinski, P. (1975), Sleep cycle oscillation: reciprocal discharge by two brainstem neuronal groups. *Science*, 189: 55-58.
Hochberg L., Bacher, D., Jarosiewicz, B. *et al.* (2012), Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature,* 485: 372-375.
Hohwy, J. (2012), Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology,* 3: 96, doi.org/10.3389/fpsyg.2012.00096
Hohwy, J. (2013), *The Predictive Mind.* New York: Oxford University Press.
Holeckova, I., Fischer, C., Giard, M.-H. *et al.* (2006), Brain responses to subject's own name uttered by a familiar voice. *Brain Research,* 1082: 142-152.
Holstege, G., Georgiadis, J., Paans, A. *et al.* (2003), Brain activation during human male ejaculation. *Journal of Neuroscience*, 23: 9185-9193.
Horikawa, T., Tamaki, M., Miyawaki, Y. şi Kamitani, Y. (2013), Neural decoding of visual imagery during sleep. *Science,* 340: 639-642.
Hsieh, P.-J. şi Tse, P. (2006), Illusory color mixing upon perceptual fading and filling-in does not result in „forbidden colors". *Vision Research,* 46: 2251-2258.

Hume, D. (1748), *Philosophical Essays Concerning Human Understanding.* Londra: A. Millar.

Hurley, M., Dennett, D. şi Adams, R. (2011), *Inside Jokes: Using Humor to Reverse-Engineer the Mind.* Cambridge, MA: MIT Press.

Ingvar, D. (1985), „Memory of the future”: an essay on the temporal organization of conscious awareness. *Human Neurobiology,* 4: 127-136.

Jackson, F. (1982), Epiphenomenal qualia. *Philosophical Quarterly,* 32: 127-136.

Jackson, F. (1995), Postscript on „What Mary Didn’t Know”. În P. Moser şi J. Trout (eds.), *Contemporary Materialism.* Londra: Routledge, pp. 184-189.

Jaspers, K. (1963), *General Psychopathology.* Chicago: University of Chicago Press.

Jaynes, E. (1957), Information theory and statistical mechanics. *Physical Review,* 106: 620-630.

Jouvet, M. (1965), Paradoxical sleep: a study of its nature and mechanisms. *Progress in Brain Research,* 18: 20-62.

Joyce, J. (2008), Bayes’ theorem. *Stanford Encyclopedia of Philosophy*.

Julesz, B. (1971), *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press.

Kandel, E. (1998), A new intellectual framework for psychiatry. *American Journal of Psychiatry,* 155: 457-469.

Kandel, E. (1999), Biology and the future of psychoanalysis: a new intellectual framework for

psychiatry revisited. *American Journal of Psychiatry*, 156: 505-524.
Kant, I. (1790), Kritik der Urteilskraft. *Kants gesammelte Schriften,* 5. Berlin: Walter de Gruyter.
Kaplan-Solms, K. şi Solms, M. (2000), *Clinical Studies in NeuroPsychoanalysis: Introduction to a Depth Neuropsychology.* Londra: Karnac.
Kihlstrom, J. (1996), Perception without awareness of what is perceived, learning without awareness of what is learned. În M. Velmans (ed.), *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews.* Londra: Routledge, pp. 23-46.
Knill, J. şi Pouget, A. (2004), The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences,* 27: 712-719.
Koch, C. (2004), *The Quest for Consciousness: A Neurobiological Approach.* Englewood, CO: Roberts and Company.
Kopelman, M., Bajo, A. şi Fotopoulou, A. (2015), Confabulation: memory deficits and neuroscientific aspects. În J. Wright (ed.), *International Encyclopedia of Social and Behavioral Sciences.* New York: Elsevier.
Krahe, C., Springer, A., Weinman, J. şi Fotopoulou, A. (2013), The social modulation of pain: others aş predictive signals of salience – a systematic review. *Frontiers in Human Neuroscience,* 7: 386.
Kurzweil, R. (2005), *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.
Lane, N. (2015), *The Vital Question: Why is Life the Way*

It Is? Londra: Profile.
Lavie, P., Pratt, H., Scharf, B., Peled, R. şi Brown, J. (1984), Localized pontine lesion: nearly total absence of REM sleep. *Neurology*, 34: 118-120.
LeDoux, J. (1996), *The Emotional Brain.* New York: Simon and Schuster.
LeDoux, J. (1999), Psychoanalytic theory: clues from the brain.
Neuropsychoanalysis, 1: 44-49.
LeDoux, J. şi Brown, R. (2017), A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Science,* 114: e2016-e2025.
Lee, J., Everitt, B. şi Thomas, K. (2004), Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science,* 304: 839-843.
Léna, I., Parrot, S., Deschaux, O. *et al.* (2005), Variations in extracellular levels of dopamine, noradrenaline, glutamate, and aspartate across the sleep-wake cycle in the medial prefrontal cortex and nucleus accumbens of freely moving rats. *Journal of Neuroscience Research,* 81: 891-899.
Leng, G. (2018), *The Heart of the Brain: The Hypothalamus and Its Hormones.* Cambridge, MA: MIT Press.
LeVay, S. (1993), *The Sexual Brain.* Cambridge, MA:
MIT Press.
Levine, J. (1983), Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly,* 64: 354-361.

Libet, B., Gleason, C., Wright E. şi Pearl, D. (1983), Time of conscious intention to act in relation to onset of cerebral activity (readinesspotential): the unconscious initiation of a freely voluntary act. *Brain,* 106: 623-642.
Lichtheim, L. (1885), On aphasia. *Brain*, 7: 433-484.
Liepmann, H. (1900), Das Krankheitsbild der Apraxie („motorischen Asymbolie”) auf Grund eines Falles von einseitiger Apraxie. *Monatsschrift für Psychiatrie und Neurologie*, 8: 15-44.
Lightman, A. (2018), *Searching for Stars on an Island in Maine.* New York: Pantheon.
Lin, P., Abney, K. şi Bekey, G. (2011), *Robot Ethics.* Cambridge, MA: MIT Press.
Linnman, C., Moulton, E., Barmettler, G. *et al.* (2012), Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage,* 60: 505-522.
Lisman, J. şi Buzsaki, G. (2008), A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophrenia Bulletin,* 34: 974-980.
Lissauer, H. (1890), Ein Fall von Seelenblindheit, nebst einem Beitrag zur Theorie derselben. *Archiv für Psychiatrie und Nervenkrankheiten,* 21: 222-270.
Lupyan, G. (2015), Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology,* 6: 547-569.
Lupyan, G. şi Clark, A. (2015), Words and the world: predictive coding and the language-perception-cognition interface. *Current Directions in*

Psychological Science, 24: 279-284.
Lupyan, G. şi Thompson-Schill, S. (2012), The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology – General,* 141: 170-186.
Lupyan, G. şi Ward, E. (2013), Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, 110: 14196-14201.
Malcolm-Smith, S., Koopowitz, S., Pantelis, E. şi Solms, M. (2012), Approach/avoidance in dreams. *Cognition and Consciousness*, 21: 408-412.
Man, K. şi Damasio, A. (2019), Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence,* 1: 446-452, doi.org/10.1038/s42256-019-0103-7
Maturana, H. şi Verela, F. (1972), *Autopoiesis and Cognition: The Realization of the Living.* Londra: Dordrecht.
Marc, J., şi Llinas, R. (1994), Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences,* 91: 1174811751.
Mathur, P., Lau B. şi Guo, S. (2011), Conditioned place preference behavior in zebrafish. *Nature Protocols,* 6: 338345.
Maxwell, J. (1872), *Theory of Heat.* Londra: Longmans, Green and Co.
Mazur, J.E. (2013), Basic principles of operant conditioning. *Learning and Behavior,* 7th edn. New

York: Pearson, pp. 101126.
McCarley, R. şi Hobson, J.A. (1977), The neurobiological origins of psychoanalytic dream theory. *American Journal of Psychiatry*, 134: 1211-1221.
McKeever, W. (1986), Tachistoscopic methods in neuropsychology. În H.J. Hannay (ed.), *Experimental Techniques in Human Neuropsychology.* Oxford: Oxford University Press, pp. 167-211.
Merker, B. (2007), Consciousness without cerebral cortex: a challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30: 63-68.
Mesulam, M.M. (2000), Behavioral neuroanatomy: large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. În *Principles of Behavioral and Cognitive Neurology,* 2nd edn. New York: Oxford University Press, pp. l-120.
Meynert, T. (1867), Der Bau der Gross-Hirnrinde und seine örtliche Verschiedenheiten, nebst einem pathologischanatomischen Corollarium. *Vierteljahrsschrift für Psychiatrie in ihren Beziehungen zur Morphologie und Pathologie des Central-Nervensystems, die physiologischen Psychologie, Statistik und gerichtlichen Medizin,* 1: 77-93, 119-124.
Meynert, T. (1884), *Psychiatrie: Klinik der Erkrankungen des Vorderhirns.* Viena: W. Braumuller.
Misanin, J., Miller, R. şi Lewis, D. (1968), Retrograde amnesia produced by electroconvulsive shock after

reactivation of a consolidated memory trace. *Science,* 160: 203-204.
Mohr, M. von and Fotopoulou, A. (2017), The cutaneous borders of interoception: active and social inference of pain and pleasure on the skin. În M. Tsakiris şi H. de Preester (eds.), *The Interoceptive Basis of the Mind.* Oxford: Oxford University Press.
Mongillo, G. Barak, O. şi Tsodyks, M. (2008), Synaptic theory of working memory. *Science,* 319: 1543-1546.
Montague, P., Dolan R., Friston, K. şi Dayan P. (2012), Computational psychiatry. *Trends in Cognitive Sciences,* 16: 72-80.
Moruzzi, G. şi Magoun, H. (1949), Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, 1: 455-473.
Motta, S., Carobrez, A. şi Canteras, N. (2017), The periaqueductal gray and primal emotional processing critical to influence complex defensive responses, fear learning and reward seeking. *Neuroscience and Biobehavioral Reviews,* 76: 39-47.
Moustafa, A., Sherman, S. şi Frank, M. (2008), A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia*, 46: 3144-3156.
Mulert, C., Menzinger, E., Leicht, G. *et al.* (2005), Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *International Journal of Psychophysiology,* 56: 65-80.
Munk, H. (1878), Weiteres zur Physiologie des

Sehsphäre der Grosshirnrinde. *Deutsche medizinische Wochenschrift,* 4: 533536.

Munk, H. (1881), *Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen aus den Jahren 1877-80.* Berlin: Albrecht Hirschwald.

Nader, K., Schafe, G. şi LeDoux, J. (2000), Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature,* 406: 722-726.

Nagel, T. (1974), What is it like to be a bat? *Philosophical Review,* 83: 435-450.

Newman, J. şi Baars, B. (1993), A neural attentional model for access to consciousness: a global workspace perspective. *Concepts in Neuroscience,* 4: 255-290.

Niedenthal, P. (2007), Embodying emotion. *Science,* 316: 10021005.

Nishimoto, S., Vu, A., Naselaris, T. *et al.* (2011), Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology,* 21: 1641-1646.

Nour, M. şi Carhart-Harris, R. (2017), Psychedelics and the science of self-experience. *British Journal of Psychiatry*, 210: 177-179.

Nummenmaa, L., Hari, R., Hietanen, J. şi Glerean, E. (2018), Maps of subjective feelings. *Proceedings of the National Academy of Sciences,* 115: 9198-9203.

Oakley, D. şi Halligan, P. (2017), Chasing the rainbow: the non- conscious nature of being. *Frontiers in Psychology*, 8: 1924, doi.org/

10.3389/fpsyg.2017.01924
Oberauer, K., Souza, A., Druey, M. şi Gade, M. (2013), Analogous mechanisms of selection and updating in declarative and procedural working memory: experiments and a computational model. *Cognitive Psychology,* 66: 157-211.
Okuda J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M. şi Yamadori, A. (2003), Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. *Neuroimage,* 19: 1369-1380.
Pace-Schott, E. şi Hobson, J.A. (1998), Review of „The Neuropsychology of Dreams". *Trends in Cognitive Sciences,* 2: 199-200.
Paloyelis, Y., Krahe, C., Maltezos, S., Williams, S., Howard, M. şi Fotopoulou, A. (2016), The analgesic effect of oxytocin in humans: a double-blind, placebo-controlled cross-over study using laser-evoked potentials. *Journal of Neuroendocrinology,* 28: 10.1111/jne.12347
Panksepp, J. (1974), Hypothalamic regulation of energy balance and feeding behavior. *Federation Proceedings,* 33: 1150-1165.
Panksepp, J. (1998), *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.
Panksepp, J. (2011), The basic emotional circuits of mammalian brains: do animals have affective lives? *Neuroscience and Biobehavioral Reviews,* 35: 1791-1804.

Panksepp, J. şi Biven, L. (2012), *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions.* New York: Norton.
Panksepp, J. şi Burgdorf, J. (2003), „Laughing" rats and the evolutionary antecedents of human joy. *Physiology and Behavior*, 79: 533-547.
Panksepp, J. şi Solms, M. (2012), What is neuropsychoanalysis? Clinically relevant studies of the minded brain. *Trends in Cognitive Sciences,* 16: 6-8.
Parr, T. şi Friston, K. (2018), The anatomy of inference: generative models and brain structure. *Frontiers in Computational Neuroscience,* 12: 90, doi.org/10.3389/fncom.2018.00090
Partridge, M. (1950), *Pre-Frontal Leucotomy: A Survey of 300 Cases Personally Followed for l-3 Years.* Oxford: Blackwell.
Parvizi, J. şi Damasio, A. (2003), Neuroanatomical correlates of brainstem coma. *Brain,* 126: 1524-1536.
Pasley, B., David, S., Mesgarani, N. *et al.* (2012), Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1): e1001251, doi.org/10.1371/journal.pbi0.1001251
Pellis, S. şi Pellis, V. (2009), *The Playful Brain: Venturing to the Limits of Neuroscience*. Oxford: One World.
Penfield, W. şi Jasper, H. (1954), *Epilepsy and the Functional Anatomy of the Human Brain.* Little, Brown and Co.
Petkova, V. şi Ehrsson, H. (2008), If I were you:

perceptual illusion of body swapping. *PLoS One,* 3: e3832, doi.org/10.1371/journal.pone.0003832
Pezzulo, G. (2014), Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive Affective and Behavioral Neuroscience*, 14: 902-911.
Pfaff, D. (2005), *Brain Arousal and Information Theory.* Cambridge, MA, Harvard University Press.
Picard, F. şi Friston, K. (2014), Predictions, perception, and a sense of self. *Neurology,* 83: 1112-1118.
Popper, K. (1963), *Conjectures and Refutations.* Londra: Routledge.
Qin, P., Di, H., Liu, Y. *et al.* (2010), Anterior cingulate activity and the self in disorders of consciousness. *Human Brain Mapping,* 31: 1993-2002.
Ramachandran, V.S. (1992), Filling in the blind spot. *Nature,* 356: 115.
Ramachandran, V.S. şi Gregory, R. (1991), Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350: 699-702.
Ramachandran, V.S., Gregory, R. şi Aiken, W. (1993), Perceptual fading of visual texture borders. *Vision Research,* 33: 717721.
Reggia, J. (2013), The rise of machine consciousness: studying consciousness with computational models. *Neural Networks,*
44: 112-131.
Riggs, L. şi Ratliff, F. (1951), Visual acuity and the normal tremor of the eyes. *Science,* 114: 17-18.
Roepstorff, A. şi Frith, C. (2004), What’s at the top in the topdown control of action? Script-sharing and

„top-top” control of action in cognitive experiments. *Psychological Research,* 68: 189-198.
Rolls, E. (2014), *Emotion and Decision-Making Explained.* New York: Oxford University Press.
Rolls, E. (2019), Emotion and reasoning in human decisionmaking. *Economics*, 13: l-31.
Rosenthal, D. (2005), *Consciousness and Mind.* Oxford: Oxford University Press.
Rovelli, C. (2014), *Seven Brief Lessons in Physics.* Londra: Allen Lane.
Runes, D. (1972), *Dictionary of Philosophy*. Totowa, NJ: Littlefield, Adams and Co.
Sacks, O. (1970), *Migraine.* Londra: Vintage.
Sacks, O. (1973), *Awakenings.* Londra: Duckworth [ed.rom.: *Revenirea la viaţă,* traducere din limba engleză de Mihaela Doagă. Bucureşti: Humanitas, 2012].
Sacks, O. (1984), *A Leg to Stand On.* New York: Simon and Schuster [ed.rom.: *Un picior de sprijin,* traducere din engleză şi note de Ioana Ilie. Bucureşti: Humanitas, 2013].
Sacks, O. (1985), *The Man Who Mistook His Wife for a Hat.* Londra: Duckworth [ed.rom.: *Omul care îşi confunda soţia cu o pălărie,* traducere din engleză de Dan Rădulescu. Bucureşti: Humanitas, 2005].
Schacter, D., Addis, D. şi Buckner, R. (2007), The prospective brain: remembering the past to imagine the future. *Nature Reviews Neuroscience,* 8: 657-661.
Schindler, R. (1953), Das Traumleben der Leukotomierten. *Wiener Zeitschrift fur Nervenheilkunde,* 6: 330.

Searle, J. (1980), Minds, brains, and programs. *Behavioral and Brain Sciences,* 3: 417-424.
Searle, J. (1992), *The Rediscovery of the Mind.* Cambridge, MA: MIT Press.
Searle, J. (1993), The problem of consciousness. *Social Research,* 60: 3-16.
Searle, J. (1997), *The Mystery of Consciousness.* Londra: Granta.
Seeley, W., Menon, V., Schatzberg, A. *et al.* (2007), Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience,* 27: 2349-2356.
Seth, A. (2013), Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17: 565-573.
Shallice, T. (1988), *From Neuropsychology to Mental Structure.* Cambridge: Cambridge University Press.
Shannon, C. (1948), A mathematical theory of communication. *Bell System Technical Journal,* 27: 379-423.
Shapiro, L. (2007), *Princess Elisabeth of Bohemia and Rene Descartes: The Correspondence Between Princess Elisabeth of Bohemia and Rene Descartes.* Chicago: University of Chicago Press.
Sharf, B., Moskovitz, C., Lupton, M. şi Klawans, H. (1978), Dream phenomena induced by chronic levodopa therapy. *Journal of Neural Transmission,* 43: 143-151.
Shewmon, D., Holmes, G. şi Byrne, P. (1999), Consciousness in congenitally decorticate children:

developmental vegetative state aş self-fulfilling prophecy. *Developmental Medicine and Child Neurology,* 41: 364-374.

Skinner, B.F. (1953), *Science and Human Behavior.* New York: Macmillan.

Solms, M. (1991), Summary and discussion of the paper „The neuropsychological organisation of dreaming: implications for psychoanalysis". *Bulletin of the Anna Freud Centre,* 16: 149-165.

Solms, M. (1995), New findings on the neurological organization of dreaming: implications for psychoanalysis. *Psychoanalytic Quarterly,* 64: 43-67.

Solms, M. (1996), Was sind Affekte? *Psyche,* 50: 485-522.

Solms, M. (1997a), *The Neuropsychology of Dreams: A Clinico- Anatomical Study.* Mahwah, NJ: Lawrence Erlbaum Associates.

Solms, M. (1997b), What is consciousness? *Journal of the American Psychoanalytic Association*, 45: 681-778.

Solms, M. (1998), Before and after Freud's „Project". În R. Bilder şi F. LeFever (eds.), Neuroscience of the Mind on the Centennial of Freud's Project for a Scientific Psychology. *Annals of the New York Academy of Sciences,* 843: l-10.

Solms, M. (2000a), Dreaming and REM sleep are controlled by different brain mechanisms. *Behavioral and Brain Sciences,* 23: 843-850.

Solms, M. (2000b), Freud, Luria and the clinical method. *Psychoanalysis and History,* 2: 76-109.

Solms, M. (2000c), A psychoanalytic perspective on confabulation. *Neuropsychoanalysis,* 2: 133-138.
Solms, M. (2001), The neurochemistry of dreaming: cholinergic and dopaminergic hypotheses. În E. Perry, H. Ashton şi A. Young (eds.), *The Neurochemistry of Consciousness.* New York: John Benjamins, pp. 123-131.
Solms, M. (2008), What is the „mind”? A neuro-psychoanalytical approach. În D. Dietrich, G. Fodor, G. Zucker şi D. Bruckner (eds.), *Simulating the Mind: A Technical Neuropsychoanalytical Approach.* Viena: Springer Verlag, pp. 115-122.
Solms, M. (2011), Neurobiology and the neurological basis of dreaming. În P. Montagna şi S. Chokroverty (eds.), *Handbook of Clinical Neurology,* 98 (3rd series), *Sleep Disorders,* Part 1. New York: Elsevier, pp. 519-544.
Solms, M. (2013), The conscious id. *Neuropsychoanalysis*, 14: 585.
Solms, M. (2015a), *The Feeling Brain: Selected Papers on Neuropsychoanalysis.* Londra: Karnac.
Solms, M. (2015b), Reconsolidation: turning consciousness into memory. *Behavioral and Brain Sciences,* 38: 40-41.
Solms, M. (2017a), Empathy and other minds – a neuropsychoanalytic perspective and a clinical vignette. În V. Lux şi S. Weigl (eds.), *Empathy: Epistemic Problems and Cultural-Historical Perspectives of a Cross-Disciplinary Concept.* Londra: Palgrave Macmillan, pp. 93-114.

Solms, M. (2017b), Consciousness by surprise: a neuropsychoanalytic approach to the hard problem. În R. Poznanski, J. Tuszynski şi T. Feinberg (eds.), *Biophysics of Consciousness: A Foundational Approach.* New York: World Scientific, pp. 129-148.

Solms, M. (2017c), What is „the unconscious", and where is it located in the brain? A neuropsychoanalytic perspective. *Annals of the New York Academy of Sciences,* 1406: 90-97.

Solms, M. (2018a), Review of A. Damasio, „The Strange Order of Things". *Journal of the American Psychoanalytic Association,* 66: 579-586.

Solms, M. (2018b), The scientific standing of psychoanalysis. *British Journal of Psychiatry – International,* 15: 5-8.

Solms, M. (2019a), The hard problem of consciousness and the Free Energy Principle. *Frontiers in Psychology,* 10: 2714, doi.org/10.3389/fpsyg.2018.02714

Solms, M. (2019b), Commentary on Edmund Rolls: „Emotion and reason in human decision-making". *Economics Discussion Papers,* No. 2019-45, Kiel: Institute for the World Economy.

Solms, M. (2020a), Dreams and the hard problem of consciousness. În S. Della Salla (ed.), *Encyclopedia of Behavioral Neuroscience.* New York: Oxford University Press.

Solms, M. (2020b), New project for a scientific psychology: general scheme. *Neuropsychoanalysis*, 21.

Solms, M. (in tipar), Notes on some technical terms whose translation calls for comment. În M. Solms (ed.), *Revised Standard Edition of the Complete Psychological Works of Sigmund Freud,* 24. Lanham, MD: Rowman and Littlefield.

Solms, M. şi Friston, K. (2018), How and why consciousness arises: some considerations from physics and physiology. *Journal of Consciousness Studies,* 25: 202-238.

Solms, M., Kaplan-Solms, K. şi Brown, J.W. (1996), Wilbrand's case of „mind-blindness". În C. Code, C.-W. Walesch, A.-R. Lecours şi Y. Joanette (eds.), *Classic Cases in Neuropsychology.* Hove: Erlbaum, pp. 89-110.

Solms, M., Kaplan-Solms, K., Saling, M. şi Miller, P. (1988), Inverted vision after frontal lobe disease. *Cortex,* 24: 499509.

Solms, M. şi Panksepp, J. (2010), Why depression feels bad. În E. Perry, D. Collerton, F. LeBeau şi H. Ashton (eds.), *New Horizons in the Neuroscience of Consciousness.* Amsterdam: John Benjamins, pp. 169-179.

Solms, M. şi Saling, M. (1986), On psychoanalysis and neuroscience: Freud's attitude to the localizationist tradition. *International Journal of Psychoanalysis,* 67: 397416.

Solms, M. şi Saling, M. (1990), *A Moment of Transition: Two Neuroscientific Articles by Sigmund Freud.* Londra: Karnac.

Solms, M. şi Turnbull, O. (2002), *The Brain and the*

Inner World: An Introduction to the Neuroscience of Subjective Experience. Londra: Karnac.

Solms, M. şi Turnbull, O. (2011), What is neuropsychoanalysis? *Neuropsychoanalysis*, 13: 133-145.

Solms, M. şi Zellner, M. (2012), Freudian drive theory today. În A. Fotopoulou, D. Pfaff şi M. Conway (eds.), *From the Couch to the Lab: Trends in Psychodynamic Neuroscience.* New

York: Oxford University Press, pp. 49-63.

Squire, L. (2009), The legacy of Patient HM for neuroscience. *Neuron,* 61: 6-9.

Stein, T. şi Sterzer, P. (2012), Not just another face in the crowd: detecting emotional schematic faces during continuous flash suppression. *Emotion,* 12: 988-996.

Stoerig, P. şi Barth, E. (2001), Low level phenomenal vision despite unilateral destruction of primary visual cortex. *Consciousness and Cognition,* 10: 574-587.

Strawson, G. (2006), Realistic monism – why physicalism entails panpsychism. *Journal of Consciousness Studies,* 13: 3-31.

Sulloway, F. (1979), *Freud: Biologist of the Mind.* New York: Burnett.

Szpunar, K. (2010), Episodic future thought: an emerging concept. *Perspectives on Psychological Science,* 5: 142-162.

Szpunar, K., Watson, J. şi McDermott, K. (2007), Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences,* 104: 642-647.

The Public Voice Coalition (2018), Universal guidelines for Artificial Intelligence, Draft 9, 23 October. Bruxelles: Electronic Privacy Information Center; https://thepublicvoice.org/ai-universal-guidelines
Thorndike, E. (1911), *Animal Intelligence.* New York: Macmillan.
Tononi, G. (2012), Integrated information theory of consciousness: an updated account. *Archives of Italian Biology*, 150: 56-90.
Tossani, E. (2013), The concept of mental pain. *Psychotherapy and Psychosomatics*, 82: 67-73.
Tozzi, A., Zare, M. şi Benasich, A. (2016), New perspectives on spontaneous brain activity: dynamic networks and energy matter. *Frontiers in Human Neuroscience,* 10: 247, doi.org/10.3389/fnhum.2016.00247
Tranel, D., Gullickson, G., Koch, M. şi Adolphs, R. (2006), Altered experience of emotion following bilateral amygdala damage. *Cognitive Neuropsychiatry,* 11: 219-232.
Turnbull, O. Berry, H. şi Evans, C. (2004), A positive emotional bias in confabulatory false beliefs about place. *Brain and Cognition,* 55: 490-494.
Turnbull, O., Bowman, C., Shanker, S. şi Davies, J. (2014), Emotion-based learning: insights from the Iowa Gambling Task. *Frontiers in Psychology,* 5: 162, doi.org/10.3389/fpsyg.2014.00162
Turnbull, O., Fotopoulou, A. şi Solms, M. (2014),

Anosognosia aş motivated unawareness: the „defence” hypothesis revisited. *Cortex,* 61: 18-29.
Turnbull, O., Jenkins, S. şi Rowley, M. (2004), The pleasantness of false beliefs: an emotion-based account of confabulation. *Neuropsychoanalysis*, 6: 5-16.
Turnbull, O. şi Solms, M. (2007), Awareness, desire, and false beliefs. *Cortex*, 43: 1083-1090.
Uhlhaas, P. şi Singer, W. (2010), Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience*, 11: 100-113.
Van der Westhuizen, D., Moore, J., Solms, M., van Honk, J. (2017), Testosterone facilitates the sense of agency. *Consciousness and Cognition*, 56: 58-67.
Van der Westhuizen, D. şi Solms, M. (2015), Social dominance in relation to the Affective Neuroscience Personality Scales. *Consciousness and Cognition*, 33: 90-111.
Varela, F., Thompson, E. şi Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press.
Venkatraman, A., Edlow, B. şi Immordino-Yang, M. (2017), The brainstem in emotion: a review. *Frontiers in Neuroanatomy,* 11: 15, doi.org/10.3389/fnana.2017.00015
Vertes, R. şi Kocsis, B. (1997), Brainstem and diencephaloseptohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience,* 81: 893-926.
Walker, M. (2017), *Why We Sleep.* Londra: Penguin.

Wang, X. şi Krystal, J. (2014), Computational psychiatry. *Neuron,* 84: 638-654.
Weiskrantz, L. (2009), *Blindsight: A Case Study Spanning 35 Years and New Developments.* New York: Oxford University Press.
Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgment to Calculation.* New York: W.H. Freeman.
Wernicke, C. (1874), *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis.* Breslau: M. Crohn and Weigert.
Wheeler, A. (1990), *A Journey into Gravity and Spacetime.* New York: W.H. Freeman.
White, B., Berg, D., Kan, J. *et al.* (2017), Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, 8: 14263, doi.org/10.1038/ncomms14263
Whitty, C. şi Lewin, W. (1957), Vivid daydreaming; an unusual form of confusion following anterior cingulectomy. *Brain*, 80: 72-76.
Wilbrand, H. (1887), *Die Seelenblindheit als Herderscheinung und ihre Beziehungen zur Homonymen Hemianopsie zur Alexie und Agraphie.* Wiesbaden: J.F. Bergmann.
Wilbrand H. (1892), Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund. *Deutsche Zeitschrift für Nervenheilkunde,* 2: 361.
Yang, E. Zald, D. şi Blake, R. (2007), Fearful expressions gain preferential access to awareness during

continuous flash suppression. *Emotion,* 7: 882-886.
Yovell, Y., Bar, G., Mashiah, M. *et al.* (2016), Ultra-low-dose buprenorphine aş a time-limited treatment for severe suicidal ideation: a randomized controlled trial. *American Journal of Psychiatry,* 173: 491-498.
Yu, C.K.-C. (2007), Cessation of dreaming and ventromesial frontal-region infarcts. *Neuropsychoanalysis*, 9: 83-90.
Zahavi, D. (2017), Brain, mind, world: predictive coding, neoKantianism, and transcendental idealism. *Husserl Studies,* 34: 47-61, doi.org/10.1007/s10743-017-9218-z
Zeki, S. (1993), *A Vision of the Brain.* Oxford: Blackwell.
Zellner, M., Watt, D., Solms, M. şi Panksepp, J. (2011), Affective neuroscientific and neuropsychoanalytic approaches to two intractable psychiatric problems: why depression feels so bad and what addicts really want. *Neuroscience and Biobehavioral Reviews,* 35: 2000-2008.
Zeman, A. (2001), Consciousness. *Brain*, 124: 1263-1289.
Zhou, T., Zhu, H., Fan, Z. *et al.* (2017), History of winning remodels thalamo-PFC circuit to reinforce social dominance. *Science*, 357: 162-168.

Index*

A

a doua lege a termodinamicii *vezi* lege
Abbott, A. 336-337 Abney, K. 375-376 Absher, J. 330-331 Abu-Hassan, K. 374 acetilcolină *vezi* neuromodulatori/neurotransmiţători activare 31, 39, 107-108, 134-150, 152-155, 166-167, 200, 205-206, 209-212, 217-218, 232-234, 310-311, 313-317, 322-323, 333-334, 336-338, 341-342, 355-356, 365-366
definiţia termenului 314-315, 337-338 *vezi şi* afecte; precizie; stare de veghe acurateţe *vezi* model generativ: acurateţe Adams, R. 351-352, 367-368, 372-373 Addis, D. 367 adicţie 54-55, 124-125, 296-297 adresare şi răspuns la întrebări 166-169, 174, 176-177, 180-182, 194, 220, 345-346 afazie 74-75, 329-331 *vezi şi* Meynert
afect senzorial *vezi* afecte: senzoriale

afecte 13, 39, 56-57, 106-109, 144-147, 151-152, 154-155, 197, 246, 249, 271-272, 278, 291-292, 304-305, 309, 321-323, 336-337, 254-256, 372-373
ca variabile categoriale (calităţi) 201-204, 219-221, 291-292, 309-310, 332-333

clasificări 110-114
dezinhibate 64-65, 79-81
fixate/stabilizate 152-153, 229-231, 278-279, 348-349, 356-357, 363-364, *vezi şi* cortex: stabilizează conştiinţa; energie latentă; memorie: pe termen scurt
senzoriale 105-106, 110-113, 117-118, 124-125, 151-152, 244245, 367-369; *vezi şi* dezgust; durere; surpriză
versus nevoi 107-110, 113-115, 147-148, 201, 211-214, 271272, 278, 289-290
vezi şi activare; emoţii de bază; pulsiune; sentimente
Africa de Sud 17, 32-33, 48, 119-120, 205-206 agnozie 72-75, 329-330
vezi şi Meynert Aiken, W. 359-360 Ainley, V. 358-359 Alberini, C. 365-366 Alboni, P. 332-333 alegere 89-91, 108-110, 121-122, 128, 150-151, 154-155, 211-212, 251-252, 273, 309-310
vezi şi valori; voluntar versus involuntar alte minţi *vezi* filosofie: problema altor minţi amânare 90-91, 202, 229-231, 278-279 amigdală *vezi* nuclei şi alte structuri cerebrale amnezie 48-49, 51-54, 75-76, 90-91, 150-151
vezi şi confabulaţie; Meynert
amplificare postsinaptică *vezi* precizie
amplitudine *vezi* precizie
anesteziologie 68-69, 77, 132-134
animale nonumane 14, 111-113, 118-119, 243-244, 292, 294-299, 303-305, 328-329
vezi şi mamifere; oi; păsări; primate; vertebrate
animale supuse intervenţiei de decorticare cerebrală 63-65, 7480, 84-85, 93-94, 148-150, 236-237, 340-

341, 366-367
anosognozie 54-55
antropomorfism 111-112 apercepţie *vezi* empirism
definiţia termenului 329
apraxie 74-75, 329-330
vezi şi Meynert
aria tegmentală ventrală *vezi* nuclei şi alte structuri cerebrale Aserinsky, E. 26-28, 321-322
Ashby, W.R. 170-171, 348-349
Astor, R. 319
atenţie 113-114, 141-142, 148-149, 153-155, 213-214, 316, 355356, 359-363, 365-368
Atkinson, R. 370-371
automatism 90-91, 108-110, 117-118, 121-122, 142-143, 147-148, 154-155, 189-190, 205-206, 229-231, 236-237, 263-264, 336-337 ca starea ideală de funcţionare cognitivă 122-123, 230-236, 310-311
autovalidare *vezi* sisteme cu autoorganizare
Axmacher, N. 319
Baars, B. 92-94, 255-256, 330-332
Bailey, P. 336-337
Bajo, A. 326-327
Bandler, R. 336-337
Baragwanath, spital, Soweto 48, 102
Barak, O. 370-371
Bargh, J. 90, 270-271, 331-332, 363
Barkay, P. 319
Barrett, L. Feldman 112-113, 122-123, 245, 333
Barth, E. 340-341
Bastos, A. 352-353
Bayes, T. 181-183

regula 190-191, 194-196, 205-208, 228-229, 351-352, 354-355, 357-358
teorema 182-184, 351-352
vezi şi probabilitate: posterioară/anterioară
Bechtel, W. 324-325
behaviorism 19-24, 26-29, 31-32, 96-99, 201-203, 293-294, 323
324, 364-365
„cutia neagră” 20-22, 98-99
recompensă şi pedeapsă 21-22, 37- 38, 48-49, 53-54, 96-99, 293-294, 323-324, 364-365; *vezi şi* eroare de predicţie a recompensei
teoria învăţării 20-24, 96-99
vezi şi lege: a efectului
Bekey, G. 375-376
Benasich, A. 346, 352-353
Benson, D. 330-331
Bentley, B. 340-341
Berlin, H. 327-329
Berridge, K. 37-38, 323-324
Berry, H. 326-327
Besharati, S. 326-327
Bienenstock, E. 364-365
bit *vezi* informaţie: biţi
Biven, L. 333-334, 337-338, 341-342, 359-360, 367-368
Blake, R. 361-362
Blake, Y. 333-334
Blakemore, S. 358-359
Blignaut, J. 361-362, 372-373
Block, N. 330-331
Blomstedt, P. 336-337
Bogen, J. 324-325, 331-332
Bohr, N. 345-346

Bouton, C. 374
Bowlby, J. 123-124, 334
Braun, A. 38-39, 69-70, 323-326
Breakspear, M. 359-360
Brentano, F. 348-349
briciul lui Occam *vezi* lege: (principiul) parcimoniei
Broca, P.P. 24-25, 329-330
Brooks, S. 319
Brown, H. 359
Brown, J. 326-327, 329-330
Brown, R. 330-331
Brucke, E. von 41-42, 307
Buckner, R. 367-368
Burgdorf, J. 335
Buzsaki, G. 359
Byrne, P. 327-328
calculatoare 21-23, 93-94, 137-138, 254-255, 267-268, 283-284, 289-290, 293-294, 298-299, 347-348, 375-376
calitatea de a fi incognoscibil 172-176, 180-181, 194, 215-216, 274, 372-373
vezi şi zonă Markov
Cameron-Dow, C. 325-326
Campbell, A.W. 71-72, 328-329
canal versus stare, funcţii 137-140, 203-204, 314-315, 336-337 Cannon, W. 314
Canteras, N. 338-339
cantitate versus calitate *vezi* filosofie: qualia
capacitate de răspuns 62-65, 67-69, 79-81, 134-135, 142-144, 313315, 337-338
vezi şi scala de comă Glasgow; stare de veghe vegetativă
Capogrosso, M. 281-283, 374

Carhart-Harris, R. 158-160, 300-301, 342-344, 359-360, 376
Carobrez, A. 338-339
Carvalho, G. 327-328, 339-340
cazuri clinice
22 ale lui Solms 83-85, 330-331
B. 57, 79-80, 105-106
domnul S. 47-55, 90-91, 96-97, 150-151, 153-154, 218-219, 239240, 256-257, 273-274, 277, 340-341
fetiţa cu hidroanencefalie 62-68, 94-95
fiica unui coleg psihiatru 68-69
Fräulein G. 72-75
H.M. 90-91
pacientul lui Blomstedt 123-124, 132, 336-337
pacientul lui Claparède 90-91, 94-97, 121-122
pacientul lui Dollman 225-228
pacientul lui Sperry 90-92, 95-96
Phineas Gage 80-81
S. M. 119-120, 333-334
T. N. 65-67
unu al lui Whitty şi Lewin 83-84
W. 80-83, 94-95
W.B. 225-226
CĂUTARE *vezi* emoţii de bază
ceva cum e *vezi* filosofie
Chabris, C. 355-356
Chalmers, D. 67, 247-273, 278-283, 285-288, 308-309, 325-326,
331-332, 345-346, 368-374
principiul coerenţei structurale 262- 265, 278-279, 281-282, 370-371

principiul dublului aspect 265-269, 278-279, 281-282, 346
principiul invarianţei organizatorice 279-282
Changeux, J.P. 330-331, 359-360
Charcot, J.-M. 21-22, 329-330
Chartrand, T. 89-90, 270-271, 321-322, 363-364
Chew, Y. 340-341
chimie 178-179
vezi şi neuromodulatori/neurotransmiţători
cibernetică 169-171; *vezi şi* feedback; semnale de eroare ciclu percepţie/acţiune 148-152, 154, 172-173, 194-195, 215-217 circuite
dopaminergice mezocorticale-mezolimbice 36-38, 53-54, 325326; *vezi şi* emoţii de bază: CĂUTARE
nervul optic 48, 66-67, 147-148, 150-151, 224-225, 327-328; *vezi şi* orbire: periferică
sistemul limbic (în general) 25-26, 192, 337-338, 367-368; *vezi şi* nuclei şi alte structuri cerebrale
Cisek, P. 359
civilizaţie 116-117, 243-244
Claparède, E. 90-91, 94-96, 121-122, 331-332
Clark, A. 238-239, 242-243, 340-341, 351-356, 360-361, 366-368 Coenen, A. 67, 327-328
Coenen, V. 333-334
cognitivism/revoluţie cognitivă 21- 24, 26-27, 56-57, 76-77, 85,
96-97, 131, 167-168, 254-257, 325-326 cogniţie *vezi şi* gândire; limbaj; percepţie
conştientă 70-71, 80-81, 83-84, 92-94, 96-97, 117-118, 142-143, 152-155, 212-213, 229-234, 238-239, 244-245, 248, 254-257, 270-272, 363-364,

365-366; *vezi şi* afecte: fixate/stabilizate; memorie: pe termen scurt
inconştientă 14-15, 43, 76-77, 80- 81, 87-96, 121-122, 154-155, 229- 232, 235-236, 251-252, 270-272, 277, 288-289, 308-309, 328-329, 363-364, 365-366
versus afect 14-15, 51-52, 56-57, 76, 80-81, 94-96, 106-107, 121123, 128-129, 142-144, 146-149, 151-155, 160, 197-198, 211212, 223, 229-230, 235-236, 244-245, 249, 271-273, 278-280, 294-295
Colby, K. 375
Cole, S. 326-327
coliculi superiori *vezi* nuclei şi alte structuri cerebrale
Collinger, J. 374
Coltheart, M. 326-327
comă 61, 63-65, 67-69, 74-76, 131, 134-135, 143-144, 160-162, 166
167, 285-286, 309, 372-373
complexitate *vezi* model generativ: complexitate
Conant, R. 348
confabulaţie 17-18, 48-55, 96-97, 150-151, 153-154, 256-258 conferinţe despre ştiinţa conştiinţei 47-38, 307, 325-326, 333-334 consolidare/reconsolidare 114-115, 140-141, 193-194, 228-236,
238-239, 252-254, 310-311, 363-366
calitatea de a fi greu de uitat a memoriei nondeclarative 121123, 193-194, 235-236
vezi şi automatism; context
conştiinţă *vezi şi* cogniţie: conştientă; percepţie: conştientă; sentimente; stare de veghe
accesare 78, 330-331; *vezi şi* sine

artificială 279-282, 285-303, 311-312, 375-376
ca travaliu de predicţie în curs 231-232, 238-239, 310-311,
360-362; *vezi şi* cogniţie: conştientă
depinde de activarea trunchiului cerebral 131-137, 233-234, 300-301, 308-309, 324-325; *vezi şi* activare
este endogenă 57-58, 132-134, 137-138, 141-142, 151-152, 197
198, 217-218, 234-235, 304-305, 366-367, 371-372
este fundamental afectivă 132-134, 151-152, 154-155, 197
198, 220-221, 271-272, 286-289, 296-297, 300-301, 308-309; *vezi şi* afecte
este incertitudine resimţită 212-213, 229-235, 236-237, 245, 291-292; *vezi şi* incertitudine; sentimente
este o formă de optimizare a preciziei 215-217, 229-230, 245, 294-295, 300-301, 356-359; *vezi şi* precizie
experienţă (conţinutul, calitatea) versus stare de veghe (nivelul, cantitatea) 26-27, 30-31, 39, 66-70, 76-78, 91-92, 128-129, 133-135, 308-309
fenomenală (primară) versus reflexivă (secundară, gânduri de ordin superior) 78, 80-81, 330-331
fizica *vezi* fizica
origine 215-216, 218-221, 250, 252-253, 259-260, 262-266, 269272, 274-275, 277, 286-288, 292, 308, 310-311
stari de ~ modificate 217-218
teorii 12, 92-94, 167-168, 255-259, 262-265, 278-

279; *vezi şi* eroarea privitoare la cortex
zori 13, 110-111, 220-221, 269-271, 278-279, 304-305 conştiinţa artificiala *vezi* conştiinţa: artificiala
context 108-109, 121-123, 145-147, 151-153, 189, 202-203, 290292, 336-337, 341-342, 354-355
preconizat 149-154, 203-206, 209-212, 236-237, 239-240, 244246, 309-310
vezi şi percepţie: contextualizeaza afectul control respirator *vezi* funcţii autonome/vegetative Conway, M. 326-327
Cooper, L. 364-365
Coren, S. 365-366
Corlett, P. 359-360
cortex 13-15, 25-26, 29-35, 39, 56-59, 62-72, 74-78, 84-85, 87-89, 118-119, 131, 134-140, 143-144, 146-149, 151-153, 159-160, 165-166, 195-196, 223-234, 236-240, 243-245, 249-250, 260-261, 263-264, 273, 278, 281-283, 285-287, 294-295, 303-304, 308309, 322-325, 328-333, 336-339, 341-344, 348-349, 352-353, 363-364
cingulat anterior 78, 81-85, 123-124, 334-336
frontal 25-26, 34-36, 39, 78-83, 92-95, 116-117, 225-227, 263264, 282-283
funcţionarea inconştienta 91-92, 136-138, 233-234, 273, 308309, 324-325, 368-369
hipocamp 25-26, 90-91, 121-122, 239-240, 282-283, 336-337,
341-342, 365-368
insular 57, 78-80
limbaj 24-25, 33-34, 74-75, 77-78, 81-85, 138-140, 283-284, 327328, 330-331

motor 33-34, 281-283
occipital 33-34, 65-67, 72-74, 78, 224-228, 234-235, 284-285, 359-360
parietal 34-36, 92-93, 146-147, 263-264
se specializează în incertitudine (complexitate, contexte) 230-231, 236-237, 366-367; *vezi şi* voluntar versus involuntar
stabilizează conştiinţa 152-153, 229-231, 278-279, 363-364; *vezi şi* amânare; memorie: pe termen scurt
temporal 74-75, 78, 139-140, 283-284; *vezi şi*
animale supuse intervenţiei de decorticare cerebrală; hidroanencefalie cortex frontal/prefrontal *vezi* cortex
Craig, A.D. (Bud) 56-57, 79-80, 330-331
Crick, F. 93-94, 248-252, 271-272, 307-308, 321-322, 368-369 cristale 157-159, 170-172, 342-343 crize *vezi* epilepsie
Crucianelli, L. 358-359
Qukur, T. 367-368

D

Damasio, A. 14-15, 56-58, 64-65, 79-83, 105-106, 133-134, 142-143, 161-163, 299-300, 326-328, 330-331, 336-337, 339-340, 343-344, 372-373
Damasio, H. 56-57, 326-327, 330-331
Darwin, C. 112-113, 169-170, 347-348

Davies, P. 247, 358-359, 368-369
Davis, E. 336-337
Debiec, J. 365-366

Decety, J. 358-359
Deco, G. 359-360
Dehaene, S. 92-93, 330-332, 359-360
Dement, W. 27-28, 321-322
demonul lui Maxwell 358-359
Dennett, D. 42, 367-368, 372-373
Depaulis, A. 336-337
depresie 19, 54-55, 123-125, 132-134
Descartes, R. 56-57, 252-253, 264-265, 369-370
dezgust 110-113, 151-152, 203-204, 244-245
    vezi şi afecte: senzoriale
Ditchburn, R.W. 233-235, 365-366
Dolan, R. 359-360
Dollman, A. 227, 319, 372-373
Domhoff, W. 367-368
domnul S. *vezi* cazuri clinice
dopamină *vezi* circuite;
neuromodulatori/neurotransmiţători
DORINŢĂ SEXUALĂ *vezi* emoţii de bază
Dudai, Y. 364-365
Duffy, E. 314
durere 13-14, 68-70, 95-97, 105-106, 110-111, 117-118,
    132, 134135, 145-147, 151-152, 232-323, 244-245,
    272-273, 289-291, 354-356, 338-340
    vezi şi afecte: senzoriale
ecuaţii 177-178, 216-217, 242-243, 302-303, 342-343
    acţiune 357-358
    conştiinţă 357-358
    divergenţă perceptivă 350-351
    energie (în general) 342-343
    energie liberă (Friston) 178-180, 215-216, 349, 357-

369
energie liberă (Gibbs) 178-179
energie liberă (Helmholtz) 178-180
energie liberă preconizată 356-357
eroare de predicţie 357-358
informaţie (entropie Shannon) 315-316
optimizarea percepţiei 216-217
percepţie 357-358
surprindere 350-351
vezi şi Bayes: teorema
Edelman, G. 92-93, 331-332
Edlow, B. 322-323, 337-339
EEG *vezi* electroencefalogramă
Ehrsson, H. 368-369
Einstein, A. 162-164, 203, 321-322, 343-344
Eisenberger, N. 335
Ekman, P. 333
electrocorticogramă 284-285
electroencefalogramă (EEG) 23-24, 32-33, 134-136, 165-167, 283
284, 313-314
desincronizată 134-135, 165-167
sincronizată 93-94, 134-135, 165-167, 258, 263-264, 341-342 undele alfa 134-136
undele beta 134-136, 352-353
undele delta 134-137
undele gama 93-94, 134-137, 258, 263-264, 352-353, 355-356 undele teta 134-136, 341-342; *vezi şi* entropie: EEG
Elisabeta a Boemiei, prinţesă 369-370 Ellis, G. 319, 372-373 emoţie inconştientă *vezi* sentiment: versus „emoţie inconştientă”

emoţii de bază 111-113, 134-137, 141-142, 152-153, 243-244; *vezi şi* afecte: clasificări; emoţii secundare
CĂUTARE 37-38, 46-48, 53-54, 57-58, 115-117, 123-126, 132134, 137-140, 145-148, 205, 230-231, 238-239, 310-311, 323326, 333-334, 341-342, 345-346, 352-353, 355-356, 365-366
DORINŢĂ SEXUALĂ 64-65, 113-116, 124-125, 127, 144-146, 212-213, 244-245, 296-297, 333-334, 339-340
FRICĂ 98-99, 105-107, 113-114, 117-123, 132-134, 137-140, 145-146, 160-162, 199, 204-206, 231-233, 235-236, 244, 290291, 303- 305, 330-334, 338-340, 368-369
FURIE 64-65, 79-80, 116-119, 122-123, 127, 132-134, 137-138, 145-146, 160-161, 199, 244-245, 304-305, 333-334, 338-340
GRIJĂ 64-65, 79, 124-126, 141-142, 145-146, 303-305, 335, 339340
JOACĂ 17-18, 62, 64-65, 79-80, 125-128, 243-244, 334-336
PANICĂ/MÂHNIRE 52-53, 98- 99, 105-107, 122-126, 132-134, 137-138, 145-147, 160-162, 203-204, 235-236, 244-246, 290291, 303-304, 307, 334-336
emoţii secundare 112-113, 122-123 empatie 127, 219-220, 243-244, 297-298, 335, 359-360 *vezi şi* emoţii de bază: JOACĂ
empirism 19-20, 61-62, 87, 96-97, 133-134, 244-245, 258, 292-293, 328-329
apercepţie versus senzaţie brută 70-73, 75-76, 78, 87-88, 90-91, 152-153, 194, 328-330

tabula rasa 70-73
vezi şi Meynert; ştiinţă: reguli
encefalită
cu *herpes simplex* 79
letargică 24-25, 217-218
energie latentă 178-179, 289-290, 348-349, 356-357, 363-264 *vezi şi* energie liberă; psihanaliză: procese secundare
energie liberă 12-13, 158-160, 183-185, 199-208, 212-220, 237238, 241-242, 289-292, 309-310, 342-343, 348-352, 356-358, 363-364
definiţiile termenului 178-180, 348-349
principiu 158-159, 162-163, 177-182, 184-185, 197, 200, 203208, 212-213, 215-219, 230-231, 237-238, 257, 288-289, 300301, 342-344, 352-353, 366-367
vezi şi ecuaţii
energie vitală 42, 45-46, 307, 324-325
England, J. 347-348
entropie 155, 162-172, 178-181, 187, 219-220, 269-270, 273-274, 286-289, 310-311, 315-316, 343-351
ca informaţie medie 165-166, 168-169, 178-180, 201-202, 286287, 315-316, 349, 266-267
EEG 165-167
epilepsie 62, 77-78, 83-84, 90-91, 364-365
eroare de predicţie a recompensei 231-233
eroarea privitoare la cortex 13-15, 57-89, 91-93, 131, 198-199, 223, 229-230, 273, 353-354
Ethier, C. 374
etică 13-14, 64-65, 68-69, 90-91, 111-112, 288-289, 396-397, 311312, 376-377

eu *vezi* psihanaliză: ego; sine
Evans, C. 326-327
„Eve Periaqueduct”187-193, 195-196, 203-204, 208-211, 288-289 Everitt, B. 365-366
evoluţie *vezi* selecţie naturală
experienţă estetică 245
vezi şi glume
explozie combinatorie 202-203, 210-211
vezi şi model generativ: complexitate
exterocepţie 113-114, 146-147, 151-155, 197-198, 206-208, 212
214, 244-246, 309-311, 353-354, 357-358, 366-367, 371-372 *vezi şi* context; percepţie; incertitudine: aplicată
Ezra, M. 337-338

F

feedback 147-150, 169-171, 173-174, 188, 190-191, 193-194, 197198
vezi şi semnale de eroare
Feldman, H. 354-359
Ferrarelli, F. 358-359
Feynman, R. 374
filosofie 13, 21, 96-97, 107, 146, 155, 252-253, 259, 272, 302-303, 333
argumentul cunoaşterii 250, 272, 369-370, 373
ceva cum e 25-26, 67-68, 77, 95, 200, 218-221, 248, 251, 256, 267-268, 273-274, 295, 308, 311; *vezi şi* qualia
dualism 218-219, 252, 264, 372-373 epifenomenalism

251-252, 370; *vezi şi* liber-arbitru inversarea qualia *vezi* filosofie: spectre inversate

metafizică 15, 19, 22-23, 25, 28-29, 46, 155, 253, 258
monism al aspectului dual 146, 155, 161-162, 200, 213, 253254, 259-260, 262-263, 265-266, 280-281, 308, 311, 372
„nimeni acasă” 67, 285-286; *vezi şi* ceva cum e; filosofie: zombi
panpsihism 94, 267-269, 272, 280-282
pe cine vizează intenţia 152, 213, 337, 348
problema altor minţi 20-21, 69-70, 77, 84-85, 174, 292-297, 348
problema fixării 92-94
problema minte-corp 19, 22-23, 25, 28-29, 44, 155, 201, 247252, 259-260, 294, 308, 369-370; *vezi şi* problema dificilă
qualia 11, 78, 201, 204, 219-221, 224, 244-246, 248, 251-252, 292, 348, 371, 373; *vezi şi* conştiinţă; filosofie: ceva cum e; percepţie: conştientă
reducţionism 249, 252, 262-264, 307, 371
spectre inversate 107-108, 244, 251, 272, 373
teleologie 169-170, 176, 219-220, 347
vid epistemic/explicativ 250, 253, 257-260, 263, 272-273, 370 zombi 67, 69-70, 231, 233, 235, 272, 293-294, 354, 370
vezi şi empirism; etică; funcţionalism; problema dificilă; ştiinţă, reguli
Fischer, D. 131, 335-336
Fischmann, T. 325-326 FitzGerald, T. 358-359 fizică 15, 42, 45-46, 155, 160, 165, 167-168, 200, 262, 265-266,

279, 287-288, 332, 344-346
a conştiinţei 162-163, 174, 198-221, 224, 247, 265-266, 269270, 275, 287-288, 311
definiţia termenului 342-343
vieţii 13, 159, 162-163, 168-173, 269-270, 274-275, 288-289, 311-312, 347-348
vezi şi entropie; materie; termodinamică; travaliu
Flechsig, P. 328-329
Fletcher, P. 359-360
Fliess, W. 45, 324-326, 321-322, 355-356, 364-365
Fonagy, P. 326-327
Forrester, G. 335-336
forţă a semnalului *vezi* precizie
Fotopoulou, A. (Katerina) 300-301, 326-327, 356-359
Frank, J. 323-324
Frank, M. 358-359
Freeman, W. 35-36
Freud, S. 14-15, 20-21, 25-26, 28-32, 38-39, 41-48, 53-59, 63- 64, 72-73, 87-91, 95-98, 102, 106-108, 115-116, 145-146, 158-161, 184-185, 197-198, 214-215, 223, 229-232, 243-244, 270-271, 287-288, 307-311, 321-332, 337-338, 340-344, 348-349, 355-357, 361-365
ideea că „conştiinţa ia naştere acolo unde se sfârseste urma mnezică” 229-232, 310-311, 363-364
ideea de „cauză principală a mecanismului psihic” 44-45, 57, 88- 89
ideea de conştiinţă în „straturile intime” ale creierului 58-59, 145-146, 304-305
ideea despre prozencefal ca „ganglion simpatic” 44-

45, 88-90, 197-198
Interpretarea viselor (1900) 28-29, 41-42
„Proiect de psihologie” (1895) 14- 15, 41-45, 159-160, 214-215,
307, 324-325, 355-357, 363-364
Zur Auffassung der Aphasien (1891) 44, 87-90, 329-331, 364-
365
vezi şi psihanaliză
FRICĂ *vezi* emoţii de bază
Friston, K. 12, 14-15, 157-163, 170-171, 174-185, 191-193, 214-216, 257-258, 273-274, 287-288, 300-301, 319, 340-343, 347-349, 351-360, 366-367, 371-372, 376-377
supa primordială simulată 157-159, 170-176, 180-182, 273274
Frith, C. 242-243, 256-257, 367-368
funcţii autonome/vegetative 61, 65-67, 107-109, 142-143
ciclul somn – stare de veghe 143-144, 313, 327-329, 336-337 controlul defecaţiei 119-120
controlul respirator 104-110, 119-120, 144-147, 160-162, 168170, 179-180, 195-198, 203-204, 304-305, 332-333, 338-340
controlul urinării 108-109, 147-148, 205-206, 332-333 digestie 108-109
echilibru energetic 13, 98-99, 105-109, 160-162, 183-184, 200203, 289-290, 352-353
echilibrul fluidelor 107-108, 146-147, 160-162, 201-203 presiune arterială 107-109, 145-146, 338-339
reflex vasovagal 104-106, 119-120, 122-124 ritm

cardiac 119-120, 144-145, 332-333, 338-339
termoreglare 105-106, 108-109, 160-163, 179-180, 191, 197198, 201, 220-221, 289-291, 294-295
vezi şi somn
funcţii vegetative *vezi* funcţii autonome/vegetative
funcţionalism 22-23, 43-44, 253-263, 273, 278-279, 308, 324-325, 370-371
vezi şi conştiinţă: teorii
FURIE *vezi* emoţii de bază
Galilei, Galileo 342-343
Galin, D. 331-332
Gallant, J. 284-285
Gallese, V. 56-57, 326-327
gama/teoria sincronizării sub 40 Hz *vezi* conştiinţă: teorii; *vezi şi* electroencefalogramă
ganglioni bazali *vezi* nuclei şi alte structuri cerebrale
Garcia-Rill, E. 336-337
Gates, W. (Bill) 375-376
gândire 117-118, 121-122, 128, 236-237, 239-244, 277, 284-286, 333-334, 374
vezi şi hoinăreala minţii
„generalizabilitate” *vezi* consolidare/reconsolidare: calitatea de a fi greu de uitat a memoriei nondeclarative; model generativ: „generalizabilitate”
Geschwind, N. 76-77, 341-342
Gilbert, J. 375-376
Ginsborg, B. 365-366
girusul cingular anterior *vezi* cortex
Gloor, P. 330-331
glume 245, 367-369
vezi şi experienţă estetică

Golaszewski, S. 336-337
Goltz, F. 75-76
Goodglass, H. 330-331
Gosseries, O. 345-346
Green, A. 326-327
Gregory, R. 353-354, 359-360
GRIJĂ *vezi* emoţii de bază
Groddeck, G. 324-325
Groote Schuur, spital, Cape Town 17-18 Grubrich-Simitis, I. 326-327
Guo, S. 375-376

H

Haikonen, P. 375-376
Halligan, P. 369-370
Hallmann, E. 324-325 halucinaţie 28-29, 36-37, 74-75
vezi şi percepţie: ca halucinaţie sau fantezie controlată Harding, D. 340-341
Harlow, J. 330-331
Hartmann, E. 323-324
Hassin, R. 363-364
Havlik, M. 370-371, 376-377
Hawking, S. 343-344, 375-376
Hebb, D. 313-314, 364-366
Heever, D. van den 319, 372-373
Helmholtz, H. von 178-179, 194, 227-228, 258, 307, 311-312, 342
344, 348-349, 359-360, 365-366
hemoragie subarahnoidă 81-84, 322-323
Herff, C. 284-285, 373-374 Hering, E. 340-341

Hesselmann, G. 355-356
hidroanencefalie 61-70, 75-78, 84-85, 93-95, 148-150, 236-237,
294-297, 327-329, 340-341, 366-367 hipocamp *vezi* cortex hipotalamus *vezi* nuclei şi alte structuri cerebrale

Hobson, J.A. 29-32, 37-39, 46-48, 322-324, 335-337, 359-360, 366
367
Hochberg, L. 374-375
Hodson, R. 372-373
Hohwy, J. 150-151, 191-193, 340-341, 352-355, 358-363, 366-367 hoinăreala minţii 236-239
vezi şi mod implicit
Holeckova, I. 336-337
Holmes, G. 327-328
Holstege, G. 336-337
homeostază 57-58, 98-99, 110-111, 160-165, 168-171, 173-174,
176-177, 179-182, 184-185, 193-194, 196-198, 200-202, 230-231,
234-235, 245-246, 286-287, 289-291, 300-301, 304-305, 308-310,
313-314, 353-354
originea 175-176, 308-309
vezi şi cibernetică; entropie; semnale de eroare; travaliu Horace, J. 370-371, 376-377
Horikawa, T. 374
hormoni *vezi* neuromodulatori/neurotransmiţători
Horwood, T. 319

Hromnik, T. 372-373
Hsieh, P.-J. 365-366
Hume, D. 70-71, 84-85, 332-333
Humphrey, N. 326-327, 333-334
Hurley, M. 367-368, 372-373

I

ierarhie predictivă 154, 187-198, 205, 207-210, 220-221, 231-233,
237, 241, 243, 245-246, 288, 310, 344, 352-354, 358-359

a elementelor anterioare 190-192, 196
imagistică prin rezonanţă magnetică funcţională (RMNf) 23-24, 132-133, 242-243, 284-285, 355-356
Immordino-Yang, M. 337-339
incertitudine 20, 99, 109-110, 112-113, 122, 150-154, 164-169, 178-181, 200, 202, 205-214, 229-236, 269-270, 275, 291-292, 295-296, 304-305, 310-311, 315-317, 351-352, 357-358, 362 aplicată 152-153, 212-213; *vezi şi* percepţie: conştientă preconizată 152-154, 207-212, 355-356 resimţită *vezi* conştiinţă: este incertitudine resimţită
speri la ce e mai bine/te pregăteşti pentru ce e mai rău *vezi* sentimente: ghidează rezolvarea problemelor
inferenţă *vezi* predicţie
informaţie 21-23, 44, 65-67, 76-78, 87-88, 136, 164-168, 178-181, 183-184, 233-234, 236-237, 255-257, 265-

270, 281-284, 286-287, 314-317, 345-349, 371-373
biţi (1/0, cifre binare) 137-140, 165-168, 346-347
byte 345
este fizică 22-23, 44, 166-168, 183-184, 265-267, 316-317, 346348; *vezi şi* materie
medie *vezi* entropie
reciprocă 180-181, 187, 352-253
informaţie reciprocă *vezi* informaţie
Ingvar, D. 328-329, 367-368
insecte 151-152, 303-304, 340-341
instinct (ghicire, intuiţie) 65-67, 121-122, 217-218, 304-305, 333334, 368-369
instinct şi reflex 107-108, 112-126, 142-143, 149-150, 191-193, 199-200, 230-231, 234-235, 309-310, 313, 335-340, 375-376 *vezi şi* automatism; funcţii autonome/vegetative
inteligenţă 13-14
vezi şi sentimente: versus inteligenţă
intenţionalitate 43, 45-46, 52-53, 62-63, 84-85, 142-144, 159-160, 169-170, 181-182, 270-271, 292-293, 295-296, 309-310, 336-338, 372-373
origine 175-176, 184-185, 197, 219-221, 229-231, 288-289, 291292
vezi şi filosofie: pe cine vizează intenţia
internet 93-94, 267-268, 272-273
interocepţie 57, 76-80, 110-111, 206-208, 217-218, 357-358 introspecţie *vezi* subiectivitate: rapoarte
introspective intuiţie *vezi* instinct
ipseitate *vezi* subiectivitate
izvor ascuns 13-14, 99, 131, 304-305

împărţire pe categorii/compartimentare a nevoilor *vezi* afecte: ca variabile categoriale
încredere *vezi* precizie
învăţare din experienţă 109-110, 112-116, 120-121, 125, 149-151, 153-154, 197, 200, 205-206, 209, 229-232, 238-239, 309-310, 336-337; *vezi şi* consolidare/reconsolidare; context; sentimente: ghidează re-zolvarea problemelor

J

Jackson, F. 369-370
Jackson, J. Hughlings 48
James, T. 319

Jasper, H. 77-78, 338-339
Jaspers, K. 330-331
Jaynes, E. Thompson 166-167, 345-346
Jenkins, S. 326-327
JOACĂ *vezi* emoţii de bază
Journal of Consciousness Studies 258, 371-372
Jouvet, M. 29-31, 322
Joyce, J. 351-352
Julesz, B. 340-341

K

Kalaska, J. 359
Kandel, E. 55-56, 326-327
Kant, I. 169-170, 342-343
vezi şi calitatea de a fi incognoscibil
Kentridge, E. 319
Kentridge, S. 319

Kernberg, O. 327
Kihlstrom, J. 90, 270-271, 332
Kinsbourne, M. 327
Kleitman, E. 26-28, 322
Knill, J. 351-352
Koch, C. 93-94, 332-333
Kocsis, B. 341-342
Kopelman, M. 326-327
Kozakova, E. 371, 376
Krahe, C. 358-359
Krikler, J. 319
Kris, E. 55
Krystal, J. 359-360
Kurzweil, R. 474

L

Lake, E. 319
Lau, B. 375
Lavie, P. 322-323
LeDoux, J. 80-81, 121-122, 326-327, 330-331, 334-335, 364-365
Lee, J. 365-366
lege
    a afectului 97-98, 108-110, 200, 309-310, 376-377
    a efectului 20-22, 97-99
    a lui Friston 184-185, 196, 206-212, 216-217, 288-289, 310-311, 376-377
    a lui Hebb 233, 365, 370
    a lui Miller 253-254, 261
    a lui Ribot 253-254, 261
    a termodinamicii, prima 162-164

a termodinamicii, a doua 162-164, 181-182, 316, 344-345, 349, 358-359
(principiul) parcimoniei 203, 237-239, 354-355
legea lui Friston *vezi* lege
Léna, I. 323-324
Leng, G. 333
leucotomie/lobotomie 34-37, 39, 83-84
Leuzinger-Bohleber, M. 326-327
LeVay, S. 333
Levine, J. 257, 369-370
Lewin, W. 83-84, 330-331

Lewis, D. 364-365
liber-arbitru *vezi şi* alegere; voluntar versus involuntar 247, 249, 251-252, 307, 351-352, 369-370
Libet, B. 252, 369-371
Lichtheim, L. 329-330
Liepmann, H. 329-330
Lightman, A. 344-345
limbaj 77, 84-85, 128, 240-244, 252-253, 256-257, 283-284, 328331, 347-348, 366-368
vezi şi cortex: limbaj; subiectivitate: rapoarte introspective
Lin, P. 375-376
Linnman, C. 337-338
Lisman, J. 358-359
Lissauer, H. 329-330
Llinas, R. 93-94, 326-327, 331-332
lobotomie *vezi* leucotomie
Locke, D. 70-71, 250-252, 272-273, 369-370, 374-375
Ludwig, K. 307

Lupyan, G. 240-242, 360-361, 367-368

M

Maclean, P. 76
Magoun, H. 131, 134, 336
Malcolm-Smith, S. 326
Malmo, R. 313-314
mamifere 13-14, 111-112, 118-119, 124-126, 292-295, 303-304, 328-329, 338-339
Marc, J. 332
Markov, A. 348

Markov, zonă 172-178, 180-181, 194-196, 215-217, 246, 269-270, 288-289, 349, 351-352
Martin, J. 319, 362, 373 „Mary” *vezi* filosofie: argumentul cunoaşterii matematică 42, 45-46, 159-160, 181-182, 184-185, 242-243, 253254, 310-311, 313-317, 342-343
materie, nu drept concept fundamental 163-165, 342-348 Mathur, P. 375
Maturana, H. 348-349 Maxwell, J.C. 358 Mazur, J. 333
McCarley, R. 322 McKeever, W. 332 memorie
ca predicţie 149-150, 152-153, 191-193, 230-233, 236-237, 239241, 291-292, 303-304, 310-311
declarativă (explicită) 43, 121-122, 191-192, 231-237, 239-240, 285-286, 303-304, 310-311, 321-322, 365-367
nondeclarativă (implicită) 43, 120-122, 147-150, 191-192, 235236, 270-271, 310-311, 321-322, 363-367; *vezi şi* consolidare/reconsolidare: calitatea de a fi greu de

uitat pe termen scurt 23-24, 34-35, 80-81, 189-194, 229-237, 253254, 291-292, 295-296, 310-311, 330-323, 363-364, 370-371 *vezi şi* consolidare/reconsolidare; învăţare
din experienţă memorie de lucru *vezi* memorie: pe termen scurt memorie declarativă *vezi* memorie
memorie emoţională *vezi* memorie: nondeclarativă; sentimente: versus la ce se referă memorie episodică *vezi* memorie: declarativă memorie nondeclarativă *vezi* memorie
memorie pe termen lung *vezi* memorie
memorie pe termen scurt *vezi* memorie
memorie procedurală *vezi* automatism: ca stare ideală de funcţionare cognitivă; memorie: nondeclarativă
Merker, B. 14-15, 61-62, 64-65, 142-143, 146-147, 149-150, 152153, 193-194, 220-221, 300-301, 327-328, 330-333, 338-342;
Mesulam, M.M. 336-337, 352-353 metafizică *vezi* filosofie metodă experimentală *vezi* ştiinţă, reguli
Meynert, T. 71-72, 75-76, 84-85, 87, 134, 329-332 asociere transcorticală 74-76, 92-93, 329-330 cortical (voluntar, mintal) versus subcortical (reflex, fizic) 71-78, 87-89, 91-92, 329-330
imagini mnezice 70-78, 87-89, 91-93, 149-152, 223, 235-237, 328-330; *vezi şi* empirism
înţelegere (cunoaştere) 70-76, 91-92, 250-251, 304-305 mitologie referitoare la creier 84-85, 330-331
mezencefal *vezi* trunchi cerebral
Miller, P. 102-104
Miller, R. 364
Misanin, J. 364
Mishra, A. 373

mod implicit 116-117, 147-148, 154-155, 205, 230-231, 238-239, 310-311, 343-344, 352-353, 365-368 *vezi şi* emoţii de bază: CĂUTAREA
model generativ 149-154, 174-175, 179-182, 187-188, 191-193, 205-206, 209-210, 213-214, 218-219, 225, 227, 230-231, 291292, 353-355
actualizare 180-183, 187-191, 193-194, 196, 205-209, 230-237, 310-311, 358-259, 363-365; *vezi şi* Bayes, T.: regula acurateţe 193-194, 203, 236-237, 349, 351-353, 366-367 complexitate 193-194, 201-203, 210-211, 236-238, 310-311, 349, 351-353, 366-367
eficienţă 236-239, 347-348, 352-353, 355-356
„generalizabilitate” 193-194, 203, 234-237, 310-311, 351-352 *vezi şi* context: preconizat; ierarhie predictivă; memorie;
predicţie: în inferenţa activă model predictiv *vezi* model generativ Modell, A. 327 modulare *vezi* activare; neuromodulare versus neurotransmisie; precizie
Mohr, M. von 359
Mongillo, G. 370 monism al aspectului dual *vezi* filosofie
Montague, P. 360
mortalitate 19, 176-177, 247, 268-269, 273-274, 277, 285-286, 291292, 301-302, 310-311
Moruzzi, G. 131, 134, 335-336
Motta, S. 338-339 Moustafa, A. 358-359
Mulert, C. 330-331
Müller, J. 42, 45-46, 208-209, 343-344
Munk, H. 72-75, 87, 329, 331-332
Munro, P. 364-365
Musk, E. 376 mutism şi akinezie 24-25, 217-218

N

Naccache, N. 92-93, 331-332

Nader, K. 364
Nagel, T. 42, 93-94, 248-249, 267-268, 308, 332, 370-371 neurochirurgie 18, 32-33, 36-37, 47-48, 51-52, 68-69, 77-78, 102
103, 260-261
neurologie 61-62, 67-68, 74-78
neuromodulare (postsinaptică) versus neurotransmisie (sinaptică) 136-142, 153-154, 217-218, 232-234, 286-287, 355356, 358-359
inhibare pe termen lung şi potenţare 139, 232-234, 316-317, 364-365
neuromodulatori/neurotransmiţători acetilcolină 30-32, 36-39, 48-49, 53-55, 140-141, 333-334, 336 337, 341-342, 355-356
aspartat 135, 137
BDNF 365-366
canabinoide 335
CCK 333-334
CRF 333-336
DBI 333-334
dopamină 24-25, 30-31, 35-39, 48- 49, 53-54, 116-117, 123-124, 132- 134, 140-141, 323-326, 335-336, 341-342
estrogen 124-125, 141-142, 333-334
GABA 135, 137
glutamat 135, 137, 333-336
histamină 140-141
LH-RH 333-334

MSH 333-334
neurotensină 333-334 noradrenalină 30-31, 132-134, 140-141, 341-342
NPY 333-334
opioide 123-124, 145-146, 335
orexină 333-334
oxitocină 124-125, 141-142, 333-336 progesteron 125-126, 141-142, 333-334 prolactină 125-126, 141-142, 335-336 serotonină 30-31, 132-134, 140-141, 341-342 substanţa P 333-334 testosteron 333-334
vasopresină 333-334 *vezi şi* tipuri de receptori
neuroni piramidali 283-284
neuropsihanaliză 46-38, 54-56
congrese 55-57, 159-160, 286-287, 330-331, 376-377
neuropsihologie
ignoră psihicul 23-25, 47-48, 53-54, 98-99, 167-168
metoda clinico-anatomică 32-33, 131-134, 143-144, 295-297
origini în empirism 70-71, 74-78, 84-85, 223; *vezi şi* empirism; Meynert
principiul dublei disocieri 33-34, 322-323
nevoi
conflictuale 117-118, 122-123, 125-126, 199, 290-291
corporale 13-14, 44-45, 57-58, 76- 77, 79-80, 88-89, 96-99, 104 111, 114-117, 142-145, 160- 161, 192, 195-198, 230-231, 352 353, 366-367; *vezi şi* funcţii autonome/vegetative
emoţionale 45, 79-80, 98-99, 106-111, 114-117, 125-126, 142

143, 160-162, 183-185, 197, 199, 205-206, 212-213, 290-291, 372-373
număr mare al 108-109, 184-185, 196, 199-204, 211-212, 219220, 273, 278, 289-290
prioritizare 108-109, 113-114, 116-118, 145-149, 190, 199-203,
205-206, 211-213, 218-219, 246, 274, 290-293, 361-362, 364365; *vezi şi* context; triunghiul de decizie din mezencefal şi oportunităţi 108-109, 113-114, 145-149, 290-292, 309-310,
361-362; *vezi şi* context; exterocepţie
vezi şi afecte: versus nevoi; senzaţii: relaţia cu nevoile Newman, J. 92-94, 255-256, 332
Newton, I. 169-170
Nishimoto, S. 284-285, 373-374
noradrenalină *vezi* neuromodulatori/neurotransmiţători
Northoff, G. 326-327
Nour, M. 359-360
nuclei şi alte structuri cerebrale
accumbens 37-38, 116-117
amigdală 103-104, 116-120, 149-150, 330-331, 333-334, 365
366
aria parafasciculară 335
aria preoptică 335
aria tegmentală ventrală 35-38, 116-117, 132-134, 140-141,
335
cerebel 149-150
coliculi superiori 65-67, 147-149, 204, 249, 287-288, 309-310,

339-340
corpul calos 224-225
corpul geniculat lateral 66-67, 150-151, 224-225, 250, 327-328 formaţiune reticulată 337-338
ganglioni bazali 148-150, 192, 270-271, 366-267
gracilis 337 hipotalamus 30-31, 57-59, 115-117, 140-142, 322-323, 337-338, 352-353
locus coeruleus complex 30-31, 132-134
nuclei parabrahiali 295-296, 335-336, 339-340
nuclei pat ai striei terminale 333-336
nucleii prozencefalului bazal 30- 31, 48, 53-54, 140-141, 322323
nucleii rafeului 30-31, 132-134, 140-141
nucleu cuneat 337
nucleu solitar 104, 337-338, 339-340
nucleu subtalamic 336-337
substanţă cenuşie periapeductală (PAG) 66-67, 115-116, 118119, 123-125, 142-148, 203-204, 211-212, 278-279, 287-288, 295-296, 309-310, 333-340, 369-370
substanţă neagră 140-141, 336-337
talamus 92-93, 263-264, 322, 324-325, 335
tegument mezopontin 30-31, 33-34, 131, 140-141, 322-323 zona locomotoare din mezencefal 147-148, 339-340; *vezi şi* triunghiul de decizie din mezencefal *vezi şi* cortex; triunghiul de de-cizie din mezencefal; trunchi cerebral
nucleii prozencefalului bazal *vezi* nuclei şi alte structuri cerebrale
Nummenmaa, L. 333

O

Oakley, D. 370
Oberauer, K. 365-366
oi 203-205, 208-209, 303-304
Okuda, J. 367
Olds, J. 76
Opatow, B. 327

opioide *vezi* neuromodulatori/neurotransmiţători
oportunităţi *vezi* nevoi Oppenheim, L. 319 orbire
    corticală 65-67, 72-74, 77-78
    periferică/subcorticală 72-75
orbire a minţii *vezi* agnozie
organisme unicelulare 13, 159 orgasm 132-135, 137
oxitocină *vezi* neuromodulatori/neurotransmiţători

P

Pace-Schott, E. 324
PAG *vezi* nuclei şi alte structuri cerebrale: substanţa cenuşie
    periapeductală
Paloyelis, Y. 359
PANICĂ/MÂHNIRE *vezi* emoţii de bază
Panksepp, J. 5, 12, 14-15, 38-39, 56-58, 65, 99, 111-113, 123-124, 131, 142-144, 147-149, 151-152, 244-245, 300, 323-324, 326328, 333-342, 360, 368
Pantelis, E. 319
Parkinson, boală 132
Parr, T. 342, 352

Partridge, M. 323
Parvizi, J. 336, 344
Pasley, B. 284, 374
păsări 111-112, 124-125, 142-143
Pellis, S. 127, 335
Pellis, V. 127, 335

Penfield, W. 77-78, 339
peptide *vezi* neuromodulatori/neurotransmiţători
percepţie 67, 70, 78, 92-93, 96-97, 148, 183-184, 195, 215-216, 223224, 253-257
a culorilor 78, 92-93, 107, 224, 227-228, 244-246, 248, 251, 257, 259-260, 265-266, 361, 368
auditivă 62, 78, 138, 152, 175, 208-209, 224, 245-246, 248, 260, 263-264, 278, 339-340, 355, 368, 372
ca halucinaţie sau fantezie controlată 151, 183-184, 237, 353354, 360-362
ca realitate virtuală 148-149, 223, 234, 237-238, 245, 339-341, 353-354, 368-369, 374-375
ca testare a ipotezei 150-151, 154, 183-184, 194, 197, 213-214, 228-229, 233-235, 353-354, 360-362
conştientă 151-154, 213, 233-242, 244-246, 248-249, 251, 254255, 257, 259-260, 265-266, 308, 354, 360-362, 366-369
contextualizează afectul 108-109, 113-114, 118, 145-147, 151154, 213-214, 220-221, 244-245, 279, 310-311, 368
determinată subiectiv 113-114, 194, 213-214, 248, 251, 259, 265-266, 353-355; *vezi şi* context: preconizat; nevoi: şi oportunităţi; predominanţă

gustativă 105-106, 113-114, 152, 224, 266-267, 361
inconştientă 62, 65-67, 76, 87, 89-92, 94-95, 148, 183-184, 191, 194, 233-235, 249, 252, 254-257, 270-272, 308, 354, 360-361, 368-369; *vezi şi* vedere oarbă
olfactivă 67, 152, 224
somatosenzorială 62, 152, 183-184, 224, 245, 278, 338-342, 368-369
tactilă *vezi* percepţie: somatosenzorială
vizuală 62, 66-67, 78, 92-93, 152, 183-184, 203, 208-209, 213-214, 224-229, 233-235, 244-246, 248-249, 251, 254-255, 257, 259-260, 263-266, 271-272, 278, 284-285, 307-308, 331-332, 339-341, 353- 355, 360, 368, 371; *vezi şi* percepţie: a culorilor; rivalitate binoculară; suprimare continuă a instantaneelor *(continuous flash suppression);* vedere inversată
vezi şi empirism: apercepţie; exterocepţie; interocepţie; psihanaliză: percepţie-conştiinţă
persoana întâi *vezi* subiectivitate: punct de vedere
perspectivă *vezi* subiectivitate: punct de vedere
PET *vezi* tomografie cu emisie de pozitroni
Petkova, V. 368
Pezzulo, G. 353, 362
Pfaff, D. 313-316, 337
Picard, F. 352
Planck, M. 345
plasticitate 140, 193, 236, 355
vezi şi consolidare/reconsolidare
plăcere/neplăcere *vezi* sentimente: valenţa

Popper, K. 29, 321-322
Porac, C. 366
Porte, J.-P. de la 41-42
Posner, M. 327
Pouget, A. 352 precizie (încredere, varianţă inversă, amplificare postsinaptică, forţă a semnalului) 138, 140, 150, 152-154, 206-221, 229-231, 234, 236, 242, 246, 273, 290-292, 300-301, 310-311, 326, 355358, 363-364, 366
ca activare/selecţie 138, 140, 151-152, 206-207, 211-212, 218,
233-234, 237, 242, 358-359
vezi şi demonul lui Maxwell; neuromodulare versus neurotransmisie predictibilitate *vezi* incertitudine; probabilitate predicţie
erori *vezi* semnale de eroare
în experimentul ştiinţific 20-21, 27-29, 37-38, 80, 99, 154, 259260, 293-295, 301-302; *vezi şi* ştiinţă: reguli
în inferenţa activă 149-154, 159, 175-176, 180-181, 194-195, 206-217, 282-283, 296, 361
în inferenţa perceptivă 183-184, 193-195, 197, 227-229, 233235, 282-283, 354, 360-362, 366-369
înnăscută 116-117, 149, 191, 196-197, 199-200, 234-235, 310, 354, 375; *vezi şi* instinct şi reflex
suprimă semnalele senzoriale 150-151, 183-184, 193, 233-235, 361-362
vezi şi Bayes: regula, teorema; incertitudine; memorie: ca predicţie; probabilitate
predominanţă 144-145, 147-149, 154-155, 197, 204-206, 211-213, 237-238, 246, 273, 337-338, 341-342,

367-368 definiţie 213-214
vezi şi nevoi: prioritizare; percepţie: determinată subiectiv
Pribram, K. 325
primate 112, 118, 142, 243, 282-284 prioritizare *vezi* nevoi: prioritizare probabilitate 164-165, 179, 181-184, 245, 315, 344, 351
posterioară/anterioară 183-184, 195, 206-208, 366; *vezi şi* Bayes: regula
problema altor minţi *vezi* filosofie
problema dificilă 12-14, 247-275, 286-288, 293-294, 299-302, 308 problema fixării *vezi* filosofie
problema minte-corp *vezi* filosofie
„Proiect de psihologie” *vezi* Freud prosopagnozie 227-228
prozencefal 29-32, 38-39, 44-45, 62-63, 88-89, 141-150, 152-154, 173-174, 197-198, 205-206, 209-210, 238-239, 309-310, 329-330, 336-340, 365-368
vezi şi cortex pseudocomă 68-69 psihanaliză 12-13, 21, 28-32, 41, 43- 45, 47-48, 53-56, 102, 106 107, 113, 123, 277
apărare 43, 356-357
cenzură 43
conţinut manifest/latent 28-29, 46
dorinţă 28-29, 31, 37, 47, 54, 58, 120, 257, 274, 326
ego 58-59, 327, 343
energie psihică 45-46, 57, 159-160, 363
inconştient 15, 43, 58-59, 72-73, 88-90, 95-96, 107-108, 270, 277, 331, 341
libido 115, 326; *vezi şi* pulsiune
metapsihologie 44-46, 57-58, 324, 341

metoda asocierii libere 28-29
percepţie-conştiinţă (P-Cs) 15, 44-45, 57-59, 64, 87-90, 95-96, 328, 331-332, 341, 357, 363
preconştient (Pcs) 88-90, 331, 363 principiul plăcerii 57, 97-98, 357 procese primare 343, 356
procese secundare 152, 343, 348, 356, 363; *vezi şi* amânare pulsiune *vezi* pulsiune
represie 43, 356, 364
rezistenţă 43, 363-364
se 57-59, 341
terapie prin vorbire 243, 293-294
testarea realităţii 54-56, 150-151, 326, 356
vezi şi Freud
psihiatrie/psihofarmacologie 29, 31- 32, 36, 54-55, 75, 83-85, 98
99, 133, 159, 218-219, 326, 330
psihoză 25, 36, 49, 132-133, 150
pulsiune 24-25, 28-29, 31-32, 37-39, 44-46, 54-55, 57-58, 76, 88-90,
106-108, 116-117, 128, 141-143, 147-148, 158-159, 184-185,
196-197, 213-214, 229-231, 271-272, 313-314, 325-326, 329-330, 337-338, 363-364
vezi şi travaliu

Q

Qin, P.331
qualia *vezi* filosofie: qualia
Queen Square, Londra 158

R

Ramachandran, V.S. 326-327, 359-360
Rance, D. 372-373
Ratliff, F. 365-366
recompensă/pedeapsă *vezi* behaviorism reconsolidare *vezi* consolidare/reconsolidare Reggia, J. 374-376
REM *vezi* somn
Remmelzwaal, L. 373 reprezentare 11, 91-92, 94-97, 99, 148-150, 174-175, 200, 217-218,

223, 322-323, 348
origine 174-175
vezi şi filosofie: pe cine vizează intenţia; Meynert: imagini mnezice
retină 193-194, 224-225, 227-228, 250-251, 257-260, 265-266, 328329, 359-360, 371-372
Reymond, E. du Bois 42, 307, 324-325
Richardson, R. 324-325
Riggs, L. 223-235, 365-366
rivalitate binoculară 227-229, 361-362
vezi şi suprimare continuă a instantaneelor *(continuous flash suppression)*
Roepstorff, A. 242-243, 367-368
Rolls, E. 37-38, 201-202, 323-324, 354-355
Rosch, E. 355-356
Rosenthal, D. 330-331
Rovelli, C. 345, 347
Rowley, M. 326-327
Royal London, spital, Whitechapel 48
Runes, D. 328-329

S

Sacks, O. 24-26, 37-38, 53-54, 56, 95-96, 98-99, 167-168, 217-218, 321, 326-327
Sacktor, T. 326-327
Saling, M. 102-103, 325-326, 330-331
saturaţie *vezi şi* homeostază; sentimente 98-99, 106-109, 114115, 230-232
scala de comă Glasgow 68-69, 80-81, 134-135, 327-328

Schacter, D. 239-241, 326-327, 367-368
Schafe, G. 364-365
Schindler, R. 224-225
Schore, A. 326-327
Schultz, W. 364-365
Schwartenbeck, P. 35
Searle, J. 250, 279-280, 282-283, 369-373
selecţie naturală 109-110, 114-115, 168-170, 290-291, 316-317 Semenza, C. 326-327
semnale de eroare 106-111, 141-143, 145-148, 150-151, 153-155, 169-171, 175-176, 183-185, 188-191, 193-198, 201, 204-212, 214-220, 234-235, 237-238, 246, 282-283, 309-311, 326-327, 351-352, 360-362, 366-367 *vezi şi* surprindere
sentimente
  ca semnale de eroare 106-111, 154-155, 160-162, 230-232, 308-309, 356-358; *vezi şi* homeostază; saturaţie; semnale de eroare
  definiţia termenului 95-96, 321-322 forţa cauzală 13-14, 96-99, 105-110, 197, 201, 212-214, 254
    256, 266-268, 270-273, 309-310; *vezi şi* alegere

ghidează rezolvarea problemelor 109-110, 112-113, 116-118, 120-121, 125-126, 128, 151-155, 200, 205-214, 229-232, 238239, 246, 309-310
importanţă 12, 14-15, 35-36, 52- 53, 80-81, 96-99, 105-110, 118119, 124-125, 128-129, 321- 322, 366-367
negate de behaviorism 19-20, 97-99, 201, 219-220, 367-368 nu sunt o funcţie a cortexului 58-59, 64-67, 69-70, 79-81; *vezi şi* eroarea privitoare la cortex
relaţia cu nevoile 44, 57-59, 76-77, 78-80, 88-89, 98-99, 104-110, 113-115, 132-134, 144-149, 184-185, 197, 200-201, 211214, 230-232, 236-237, 271-273, 309-310, 362, 366-367; *vezi şi* homeostază; nevoi: prioritizare
sunt cu necesitate conştiente 57-58, 94-96, 102, 107-109, 271274, 308-309, 321-322
sunt forma fundamentală a conştiinţei 132-134, 151-152, 154155, 197-198, 203-204, 217-218, 220-221, 246, 271-272, 285287, 308-309
sunt o funcţie a trunchiului cerebral 132-134, 144-146 valenţa 13, 57, 98-99, 104-105, 108-111, 119-120, 146-147, 157, 200-201, 203-204, 219-220, 244-245, 291-293, 296-297, 309310, 332-333, 364-365;
versus „emoţie inconştientă” 95-96, 102, 121, 271-272, 330332
versus inteligenţă 13-14, 220-221, 285-286, 288-289, 293-295, 298-299, 304-305, 330-331
versus la ce se referă 90-92, 94-96, 103-106, 121-122, 128-129, 235-236, 304-305
vezi şi afecte; conştiinţă

serotonină *vezi* neuromodulatori/neurotransmiţători
Sessions, R. 354-355
Seth, A. 358-359
Shallice, T. 324-327
Shannon, C. 166-169, 265, 315-316, 345-346
Shapiro, T. 326-327
Sharf, B. 323-324
Sherman, S. 358-359
Shewmon, D. 327-328
Shiffrin, R. 370-371
Shipp, S. 351-352
Shock, J. 319, 372-373
Simons, D. 355-356
simţuri *vezi* percepţie
sincopă vasovagală 103-105, 146-147
vezi şi funcţii autonome/vegetative: reflex vasovagal
sine 23-25, 42, 52-53, 56-57, 71-72, 78-80, 84-85, 368-369 procesare legată de sine 78, 83-84 SINELE primar *vezi* triunghiul de decizie din mezencefal *vezi şi* psihanaliză: ego; subiectivitate
Singer, W. 326-327, 358-359
sistem endocrin 140-142, 195-196
sisteme cu autoorganizare 157-159, 169-182, 190-191, 200-202, 205-206, 219-220, 230-231, 237-238, 267-268, 273-274, 288-290, 291-292, 309-311, 343, 347
care se autovalidează 180-182, 214-216, 219-221, 269-270, 288-289, 293, 311-312
definiţia sintagmei „care se autovalidează” 351-352
vezi şi homeostază
sistemul reticulat activator 39, 131-138, 140-145, 151-155, 162, 173-174, 209-212, 218-219, 232-233, 278-

279, 308-309, 322-323, 336-338, 364-366
vezi şi activare; circuite; nuclei şi alte structuri cerebrale situaţii neprevăzute *vezi* incertitudine
Skinner, B.F. 97-98, 332-333
Skotnes, P. 319
Societatea de Fizică din Berlin 42, 45-46, 178-179, 311-312, 343344
Solms, K. Kaplan 326-327, 330
Solms, L. 17-19, 101-102, 104
Solms, M. 322-336, 341-342, 344, 354-355, 357-362, 364-365, 367-369, 374-376
somn 26-31, 48, 65, 84-85, 99, 116-117, 134-140, 166-167, 199, 214-215, 238-239, 290-291, 301-302, 323-324, 336-337, 366-367 REM/non-REM 27-35, 37-38, 48, 95-96, 128, 134-135, 137-138, 322-324, 333-334, 336-337
vezi şi funcţii autonome/vegetative: ciclul somn – stare de veghe
Sperry, R. 90-92, 94-96
spike trains 139, 180-181, 224-225, 245
Squire, L. 331-332
stare de conştienţă minimă 166-167
stare de repaus *vezi* mod implicit
stare de veghe 26-27, 62, 67-68, 134, 138-140, 142-144, 314-315 *vezi şi* activare; conştiinţă: experienţă versus starea de veghe
stare de veghe vegetativă *vezi* stare vegetativă
stare vegetativă 61-65, 67-70, 75-76, 134-135, 142-143, 166-167, 285-286
Stein, T. 362
Steiner, R. 326-327

Stephan, K. 352-353, 359-360
Sterzer, P. 362
stimul/răspuns 21-24
stimulare cerebrală profundă 111-112, 132-134, 144-146, 295
297, 330-331, 336-339
stimulare optogenetică 259-260, 372
Stoerig, P. 340-341
Strachey, J. 45, 325-327
Strawson, G. 331-332 subiectivitate 11, 13, 19-21, 24-25, 33-34, 42, 44-55, 95-97, 109- 110, 174-175, 182, 201, 204, 218-219, 248, 251-252, 265-266, 279, 291-293, 296-297, 308, 310
ipseitatea 174, 176, 219-220, 292-294, 309-310, 372-373 origine 174, 218-220, 252-253, 258-260, 263-266, 274-275, 278279, 308-309
punct de vedere 11-12, 49-53, 148-149, 174, 176, 218-220, 237240, 256-257, 259-260, 265-266, 288-289, 291-292, 310, 339341, 348-349, 368-371
putere cauzală; *vezi şi* liber-arbitru 25-26, 110-111, 212-214, 250-252, 266-268, 270-272, 309-310
rapoarte introspective; *vezi şi* limbaj 20-22, 24-29, 31-32, 3435, 41-42, 77-78, 84-85, 95-97, 201, 225-227, 256-257, 293294, 328-331, 333-334, 396-370,
vezi şi adresare şi răspuns la întrebări; filosofie; sine
substanţa cenuşie periapeductală *vezi* nuclei şi alte structuri cerebrale
Sulloway, F. 325-326
supa primordială *vezi* Friston
supraajustare *vezi* model generativ: complexitate
suprimare continuă a instantaneelor *(continuons flash*

suppression) 240-242, 367-368; *vezi şi* rivalitate binoculară surprindere 179-181, 230-231, 316-317, 349-352
surpriză 110-111, 151-152, 154, 244, 367-368
vezi şi afecte: senzoriale
Szpunar, K. 367-368

Ş

ştiinţă, reguli 19-20, 29-30, 41, 68-69, 99, 258, 278, 292-293, 296297, 299-302, 321-322, 372-373
vezi şi empirism; predicţie: în ştiinţă; testarea ipotezei: în ştiinţă

T

tegment mezopontin *vezi* nuclei şi alte structuri cerebrale
teoria informaţiei integrate *vezi* conştiinţă: teoriile
teoria reintrării *vezi* conştiinţă: teorii
teoria spaţiului de lucru global *vezi* conştiinţă: teorii
terapie electroconvulsivă (ECT) 231-232, 364-365
termodinamică 159-160, 162-166, 178-180, 236-237, 274, 315-316,
345-346, 349
vezi şi lege: a termodinamicii
termoreglare *vezi* funcţii autonome/vegetative testarea ipotezelor
în percepţie 150-152, 154-155, 183-184, 194, 213-214, 228-229, 233-235, 277-278, 360-362

în ştiinţă 19-20, 26-31, 36-37, 54- 55, 64-65, 67-58, 80-81, 259
260, 277-278, 286-288, 293, 299-302; *vezi şi* ştiinţă, reguli *vezi şi* Bayes: regula
testul Turing 293-295, 375-376
Thomas, K. 365-366
Thompson, E. 355-356
Thompson-Schill, S. 367-368
Thorndike, E. 97-99, 332-333
Tinbergen, N. 313
tipuri de receptori 138, 140-142
a-MSH 333-334
p-opioid 335-336
nicotinic 31

vezi şi neurotransmiţători/neuromodulatori
tomografie cu emisie de pozitroni (PET) 23, 47, 133-134, 137-138 Tononi, G. 93-94, 331-332, 359
Tossani, E. 334-336
Tozzi, A. 346, 352-353
Tranel, D. 326-327, 330-331, 333-334
travaliu 45-46, 57-58, 88-89, 105-108, 116-117, 141-142, 146, 150, 152-153, 160-165, 181-182, 184-185, 197, 229-232, 235-239, 246, 286-287, 289-290, 297-298, 361-363
triunghiul de decizie din mezencefal 146-151, 153-154, 162, 173174, 196, 203-206, 209-210, 218-219, 309-310, 338-339, 364-365 trunchi cerebral 14-15, 29-31, 34-36, 38-39, 56-59, 65-67, 69-70, 131-137, 141-146, 193-194, 244-245, 249-250, 309, 322-323, 331-332, 336-337, 339-340, 352-353, 359-360, 261-262 bulb rahidian 337-340

canal central 142-143, 145-146, 352-354
definiţia termenului 322-323
diencefal 142-143, 193-194, 338-340; *vezi şi* nuclei şi alte structuri cerebrale: hipotalamus; talamus
mezencefal 132-134, 142-148, 278-279, 294-295, 303-304, 338340; *vezi şi* triunghiul de decizie din mezencefal
punte 29-31
vezi şi activare; nuclei şi alte structuri cerebrale; sentimente; sistemul reticulat activator
Tsakiris, M. 256-257
Tse, P. 365-366
Tsodyks, M. 370-371
Turing, A. 375-376
Turnbull, O. 293-294, 326-327, 333-334, 374-376
Turner, M. 326-327
Uhlhaas, P. 359
University of the Witwatersrand, Johannesburg 32-33, 41-42 Urbach-Wiethe, boală 119-121

V

valenţă *vezi* sentimente: valenţa
valori 13-14, 104-106, 108-109, 124, 128, 145-147, 151-152, 164165, 200-201, 210-211, 219-221, 288-292, 296-299, 303-304 succes reproductiv 113-114, 296 *vezi şi* selecţie naturală
Varela, F. 213-214, 348-349, 355-356
varianţă *vezi* precizie
vedere inversată 224-228
vedere oarbă 65-67, 77-78, 148-149, 327-328, 333-334,

340-341 Velmans, M. 326-327
Venkatraman, A. 337-338
vertebrate 111-112, 142-143, 151-152, 200, 219-221, 278, 295-296, 303-304
Vertes, R. 341-342
viaţă *vezi* fizică; homeostază; mortalitate; selecţie naturală
vise 12, 26-41, 46-49, 53-54, 73-75, 83-84, 95-96, 116-117, 120-121, 128, 133-137, 217-218, 237-239, 284-285, 322-326, 229-230, 359-360, 366-367, 374
voliţiune *vezi* intenţionalitate
voluntar versus involuntar 21-22, 61, 75-76, 87, 107-110, 113115, 117-118, 120-122, 128, 142-143, 147, 149-151, 181-182, 191-192, 200, 206, 229-230, 237-238, 277, 294-296, 310, 314315, 363

definiţia termenului „voluntar” 108-109 *vezi şi* Meynert

W

Walker, M. 336-337
Wang, X. 359-360
Ward, E. 240-241, 367-368
Watt, J. 375-376
Watts, J. 35-36
Weiskrantz, L. 66-67, 327-328
Weizenbaum, J. 375-376
Wernicke, C. 24-25, 74-75
Westhuizen, D. van der 334-336, 362, 373
Wheeler, J. 167-168, 176, 220-221, 245, 265-266, 346-

347, 371-372 White, B. 148-149, 340-341
Whitty, C. 83-84, 330-331
Widlöcher, D. 326-327
Wiener, N. 169-170
Wilbrand, H. 72-74, 328-330
Wolpert, D. 359
Wyzinski, P. 322-323

Y

Yang, E. 362
Yovell, Y. 333-334
Yu, C. Kai-Ching 323-324

Z

Zahavi, D. 371-372
Zald, D. 362
Zare, M. 346, 35-353
Zeki, S. 224, 321-322, 360
Zellner, M. 325-327
Zeman, A. 67, 134, 327-328
Zenon (din Elea) 345
zgomot *vezi* precizie
Zhou, T. 334-336, 367-368
zombi *vezi* filosofie

* Numărul paginilor se referă la ediţia tipărită a cărţii şi la ediţia digitală în format PDF.